INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 01-01-2022

Accepted: 29-04-2022

Published: 24-06-2022

Citation: Dalal D, Vasani R (2022) An Experimental Study to Locate Leakage in the Water Distribution Network using Real-time Wireless Sensor and Machine Learning Algorithm. Indian Journal of Science and Technology 15(23): 1133-1142. https://doi.org/10.17485/IJST/v15i23.2

* **Corresponding author**.

dhdalalphd7@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

An Experimental Study to Locate Leakage in the Water Distribution Network using Real-time Wireless Sensor and Machine Learning Algorithm

Dipesh Dalal[1]*, Rupesh Vasani[2]

1 Civil Engineering Department, Government Polytechnic for Girls, Gujarat Technological University, Ahmedabad, 380015, Gujarat, India
2 Civil Engineering Department, Sal Education Campus, Gujarat Technological University, Ahmedabad, 380060, India

Abstract

Objective: To pinpoint the leakage location using a machine learning algorithm on a real-time basis. **Method**: A laboratory experimental model was developed using wireless sensors for real-time data collection and monitoring the changes in the pressure and flow in presence of leakage under a different scenario. Modification in laboratory model was made considering Loop network of distribution pipes as compared to the simple experimental model. The model has been validated in EPANET software. The machine learning algorithm, along with the K-fold approach has been used to locate leakage and it was compared with other algorithms like Support Vector Machine (SVM), logistic regression, multilayer perceptron, Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM), Principal Component Analysis (PCA), Artificial Neural Network (ANN), Random Forest, Gradient Boosting Tree (GBT) Classification, and Convolutional Neural Network (CNN). **Findings**: In this research, a system was developed based on real-time data received by Wireless Sensor Network (WSN) was demonstrated and it is able to find even a small leak by monitoring pressure and flow. The K-fold approach in machine learning algorithm has been used to locate the leakage in different variations made in the experimental model in terms of pressure, leak size, cross-section of pipe, and profile level of the pipe network. A comparison was also presented with other machine learning algorithms of recent research in terms of accuracy, size of the opening for the leak, type of experimental model used, and variables considered to locate the leakage. **Novelty**: It is the first research kind of work that shows the average accuracy of 78% to locate leakage based on the real-time experimental data in a loop network using the K-Fold approach.

Keywords: Leakage; Leakage Detection; Kfold approach; Pressure Measurements; Water Distribution Network; Wireless sensor

1 Introduction

There are mainly two types of leak detection systems. static leak detection systems inform the water network management of the existence of a leak almost immediately, whereas dynamic leak detection systems have information of a leak possibility so that they can be mobilized for investigation[1]. In static leak detection systems, we are relying on sensors and data collectors that are placed within the water network which can transfer data periodically. The effective solutions before implementation of such static leak detection system are determined by different sets of hydraulic data and machine learning techniques based on a dataset using experiments or models of real networks. In this research, a comparative experimental study was presented of four different machine algorithms named k-nearest neighbors, support vector machines, logistic regression, and multilayer perception indicate that the accuracy to find the leakage in the water distribution network is dependent on many factors including the size of leakage[2]. Researchers of this study proposed a data based leak detection model for leak identification that showed a good result and has an accuracy of 90% at all points except singularities by the confusion matrix method[3]. Researchers of this study represent an investigation of the capacity of six machine learning methods presented by the data generated using EPANET software, indicate that the supervised logistic regression and random forest method performed well to localize the leakage[4]. A multi-strategy ensemble learning(MEL) was presented in this research as an effective solution for an improvement in leak detection using acoustic techniques[5]. In this study, real-time transient model has been used to avoid the large numbers of false alarms with some data mining methods having its own merits and demerits[6]. Experimental research presented by the researchers for water pipeline leakage detection based on machine learning and wireless sensor networks indicates that the simulation analysis and experimental results using Support Vector Machine (SVM) can detect leakage effectively and has lower energy consumption[7]. Many machine learning algorithms like support vector machines, logistic regression, k-nearest neighbors, multilayer perceptron, maximum entropy version of least square twin K-class support vector machine known as MLT-KSCV algorithm, random forest, gradient boosting tree classification model, confusion matrix as well as mathematical modeling were used to find the leakage points based on collected data or simulated data whose accuracy ranges between 30% to 100% are summarized below.

Table 1. Summary of literature review

Reference	Types of data field data/ Experimental data/ Simulation data	Variable considered	Leakage detection method/ Machine Learning Algorithm used	Accuracy to locate leakage	Key Points to discuss
(2)	Database taken of two networks from simulation	Flow & Pressure data	Different Algorithm K-nearest neighbor, Support Vector machine, Logistic regression & Multi-layer Perceptron	Different accuracy given for different scenario ranging from 60% to 90%	• Four algorithms compared and conclude that SVM algorithm give better result as compared to Logistic Regression (LR), Multilayer Perception (MLP) and K-nearest Neighbors (KNN). • Large leaks effect the accuracy to locate the leakage as compared to small leaks.
(3)	Experimental real measure data with linear pipe having only one branching.	Flow data	Flow meter is used to measure data and then RNN-LSTM was used for prediction with confusion matrix	90% accuracy accept at one or two leakage points.	• Only one branching with linear pipe was considered with six measurement flow points. • Accuracy varies between 99.81% to 46.46% at different points and accuracy claimed in research is expect point F2.

Continued on next page

Table 1 continued

(4)	Simulation data of different scenario was used by EPANET Software	Flow & Pressure data	Only in decision tree method flow & pressure data combined used while in other methods like logistic regression, random forest, PCA, ANN and k-method either flow or pressure data was used.	Accuracy to 98% to 100% with the consideration of only flow was achieved.	• Research is only based on Simulation, No Experimental or real data taken. • 80% data used for training and 20% data used for testing.
Reference	Types of data field data/Experimental data/Simulation data	Variable considered	Leakage detection method/Algorithm used	Accuracy	• Comparison of Six Machine Learning Methods (ML) was given and each method is used by considering only flow or pressure parameters except the decision tree method. • The campus zone divides into five zones in simulation and with the decision tree method 83% accuracy was achieved in zone 2 & 5
(5)	Real field data was collected from multiple cities of North America.	Acoustic data	Gradient Boosting Tree (GBT) classification model, ANN-PCA & KNN-PCA were used.	Accuracy is given separately for leak and no leak conditions by all three methods ranging from 62% to 96%.	• Key Points to discuss
(6)	Simulation data generated of Single pipeline for Oil & Gas pipeline.	Pressure & Flow data	Mathematical modelling with ANN, SVM and Machine Algorithm	Suggesting that ANN is poor in case of small set of samples, while use of the machine learning is better in prediction of leakage using synthetic data i.e., data generated from piping models.	• Accuracy is given individually for leak and no-leak conditions. • Researcher themselves suggested for generating more hydraulic data for further improvement.
(7)	Experimental & Simulation Based on exposed aluminum plastic composite pipe.	Flow & Pressure Data for simulation and Acoustic data for experiment work	Simulation done in OPNET Modular 14.5 version & Support Vector Machine Used	Accuracy was 98%	• Surface pipeline for oil & gas considered. • Researcher suggest ANN too much rely on training samples to achieve accuracy, SVM able to detect small leakage based on sensor data but the use of machine learning has a significant role in finding leakage.
(8)	Experimental data based on 200m linear pipe.	Acoustic Data	Least Square Twin K-class support vector machine.	Accuracy is 96.23% but algorithm may crash when data samples are very large.	• Actual experimental work based on acoustic signals while flow & pressure data only considered in simulation. • Pressure less than 0.3 Mega Pascal (MPa) is not considered. • No leak size was defined.

Continued on next page

Table 1 continued

(9)	Simulation data based used	Flow & Pressure data	Flow Master simulation was used with Transfer Learning One dimensional Convolutional Neural Network (TL1DCNN) approach	Accuracy given for different cases ranging from 74% to 90%.	Acoustic single generated from 200 m linear pipe was used as data 300 reading were taken for each different case of no leak, small leak and large leak were considered Average is different for all above three cases ranging from 93% to 98% based on only 900 readings.

The comparison of the different research works in terms of features, findings, parameters, methods, accuracy, and state of art is given in Table 1. In this perception, it is clearly understood that detecting and identifying the location of leakage is a very complex problem and it depends upon many factors. An attempt has been made to identify and locate the leakage from real-time data collected through the experimental model with different variations in this research.

2 Methods and Materials

2.1 Experimental Laboratory Model

The experimental laboratory model was made of the Unplasticized Polyvinyl Chloride (uPVC) as per American Society for Testing Materials (ASTM) standards D 1785[(10)]. The reason behind the selection of uPVC material due to its worldwide use in drinking water distribution systems nowadays. The total length of the model was kept 51 feet with two loops. The other technical details of model were presented in Table 2. An overhead tank at a height of 12 meters was kept as a source of water and to create different pressure scenarios, Variable Speed Pump (VSP) was installed which is capable to deliver a discharge of 1584 Liter Per Hour (LPH). Total eight different leakage positions were kept in the model as shown in a schematic diagram of the experimental model Figure 1. A wireless sensor box (Structure as shown in Figure 2) was also made in the laboratory to transfer the data from sensors to the server.

2.2 Calibration of a Pressure sensor

The data and results obtained from any experiments are only reliable when the equipment and all the accessories are well-calibrated before the readings. A small pressure sensor calibration model was designed and used to calibrate the pressure sensors with a pressure gauge. In this model, a pipe of 1 meter was taken and pressure gauge and pressure sensor were installed at $1/3^{rd}$ and $2/3^{rd}$ distance from the same end. Nearly 100 readings with different pressure ranging from 0 Pascal per Square Inch (PSI) to 30 PSI were taken which shows the same value in both that are pressure sensor and pressure gauge.

Table 2. Accessory used in Laboratory Network Model

Details	Description
Material	uPVC
Section of Pipe	Circular as per ASTM D 1785
Inner Diameter	12.7 mm
Outer Diameter	21.34 mm
Length of Network	51 feet, From Inlet to Outlet
Number of Loop	02 nos.
Elevation of Node	Every node kept at 2 feet elevated from Ground Level
Variable Speed Pump	01 Nos., 0.50 Horse Power (H.P)., 240 Voltage (V) single phase 2880 Revolution Per Minute (RPM), Discharge1584 LPH @ 10 m height
Water Meter	01 Nos. of 15 mm Size as per International Organization for Standardisation (ISO) 9001
Pressure Gauge	01 nos. of $\frac{1}{2}$ inch Diameter having measurement range 0-1.2MPa.
Wireless Flow Sensor	Able to work under pressure of 1-75 Mega Pascal (MPa), range 1-30 Liter Per Minute (LPM)
Wireless Pressure Sensor	Made of Carbon Steel Alloy, 5 V D.C., range 0-1.2MPa.
Leakage Positions	Total Eight different leakage positions were created in Model.

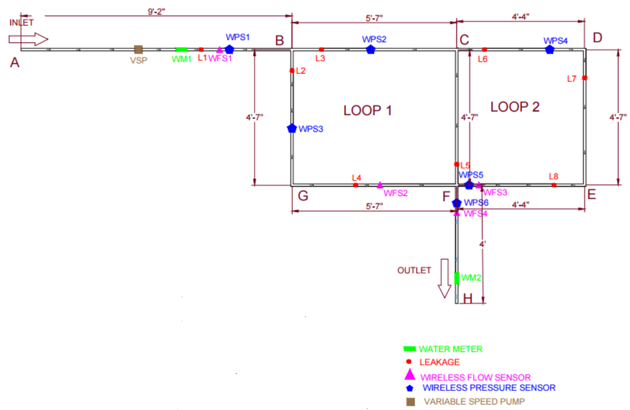

Fig 1. Schematic diagram of the Experimental Model

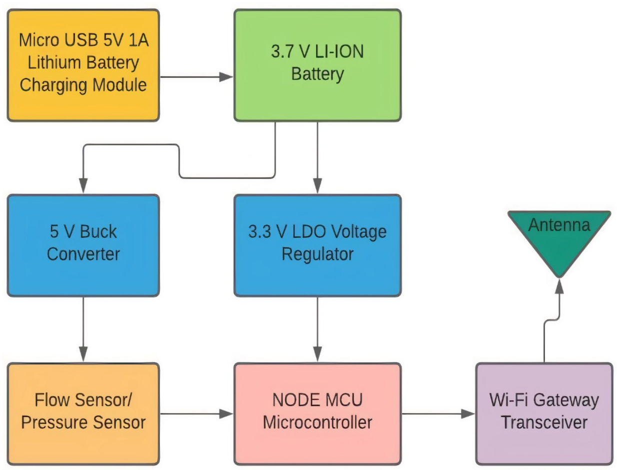

Fig 2. Structure of wireless sensor box

2.3 Experimental Work

The laboratory model was tested under different variations made in presence of eight different leakage positions. Pressure and flow sensors were just placed nearby to leakage points to record the exact changes of flow parameters. The measurement of pressure was recorded in two different units PSI and Kilopascal (KPa) and for flow, it is in LPM. The logic behind the record of pressure in two different units is that due to a small leakage very small change in pressure values and it can be reflected more easily in smaller units of Pressure that is KPa. To monitor flow parameters traditionally, one pressure gauge and two flow meters have also been installed. To match the scenario of the model with real field networks of pipe, two different experimental models of uniform and non-uniform c/s were prepared with leveled and sloppy profiles. The EPANET software was used for the validation of flow parameters. From similar research, pressure is identified as a more dominant and important factor as compared to flow to detect a small leak in the network. So initially, the data cycle for the pressure was kept as 4 seconds, while for the flow it was 30 seconds. That uneven data cycle for pressure and flow directly leads to the mismatch of recorded reading. So, it was very difficult to locate leakage using an algorithm from the huge collected uneven data. To overcome this difficulty, it was kept identical afterward as 13 seconds for pressure and flow.

Numbers of readings were recorded with the following variations keeping in mind that the location of leakage depends upon many parameters.

1. Pressure Condition: - Low, Medium, and High pressure
2. Cross-section of pipe: -Two different models made One is of the uniform pipe section, while other is of non-Uniform type section.
3. Size of Leak: - Variations made from 0% to 90% leak of pipe diameter (leak and No-leak scenario)
4. Topographical Condition: - Three different variations made concerning to slope of network i.e.

    (a) Leveled network
    (b) 1 in 25 Slope (Gentle Slope)
    (c) 1 in 15 Slope (Steep Slope)

2.4 Machine Learning Algorithm

Different machine learning algorithms and mathematical models are used to predict leaks and the location of leaks from the various data collected by many researchers. For prediction by any algorithm, it must be trained to understand the system and environment from the feed data, so that when a water leak occurs, the system may know what is happening and what to do with the received data [11,12]. Recent advancements in machine learning algorithms for the prediction of abnormal events using historical data collected from sensors for a particular water distribution network will allow finding the root cause of the problem at a specific part of the water distribution network. In this research paper, the K-fold cross-validation approach is used in machine learning algorithms for the validation of the water distribution network model as this is a popular clustering method that minimizes the errors in clustering.

In K-fold cross-validation approach, K means the number of groups so the whole data set is divided into the k number of groups for the model validation purpose. Like if K =10 then the data set is split into 10 different groups. Cross-validation techniques are most widely used for model validation because every time unseen data is represented to model for the prediction it is excluded during the training phase. Hence the k-fold validation will provide unbiased and realistic estimations of the model compared to leaving one out cross-validation, holdout validation, and other.

The following steps are followed for K-fold validation in this research.

1. Shuffle the whole database randomly
2. Split the dataset into k groups
3. For each unique group:

    (a) Take any group as a test data set
    (b) Take the remaining groups as a training data set
    (c) Fit a model on the training set and evaluate it on the test set
    (d) Retain the evaluation score and discard the model
    (e) Repeat steps 1 to 4 for each groupacy

4. Summarize the predicting accuracy obtain for each group and then take average accuracy

3 Results and Discussions

To achieve the main objectives of the research, changes in flow parameters i.e., flow & pressure were used to predict the leakage while a machine learning algorithm was used to locate the location of leakage. Sample graphs of pressure and flow under two different leakage conditions, one from each loop were presented here. Figures 3 and 4 show pressure and flow graphs with real-time stamps in presence of leakage position no.2 made in loop 1. We can see the minor and major changes in the values of pressure and flow indicating small and large leak size in the pipe. Same way, Figures 5 and 6 show the changes in pressure and flow in presence of minor and major leaks made at leakage position no. 7 in loop 2. It is observed from all graphs that when there is a leakage in any place within the network there is a change in the pressure values of all pressure sensors irrespective of the position of leakage in particular loop of the network. From the observation of flow graphs, only those value of flow sensors show change when the leakage was made in that particular loop only. The first objective to detect leakage was achieved by observing the changes in real-time graphs of pressure and flow. For the second objective to locate the leakage, we compiled all the data, shuffled it and divided this data into two different sets that is testing set and training set. K-fold approach was used in the machine learning algorithm. The flowchart of prediction of leakage with use of machine learning algorithm was shown in Figure 7. The algorithm was trained with training data set which included a compilation of all the data from different variations. Similarly, some readings were kept as testing data sets for testing the algorithm after training. A user-friendly Graphical User

Interface (GUI) was also made for the prediction of leakage and it is presented in Figure 8. It shows that for the use of GUI, we have to enter data of six pressure sensors and four flow sensors as input and as result of this we get the location of leakage as output.

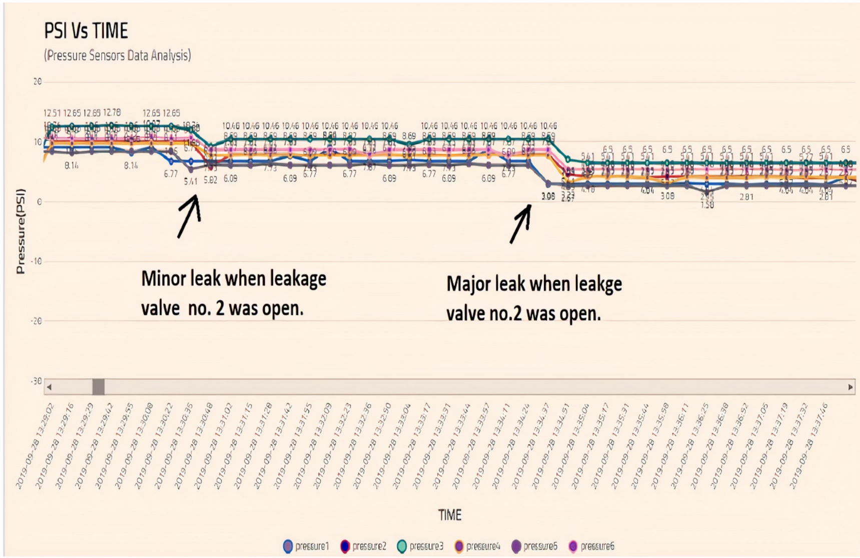

Fig 3. Pressure sensor graph when leakage valve no. 2 was open in loop 1

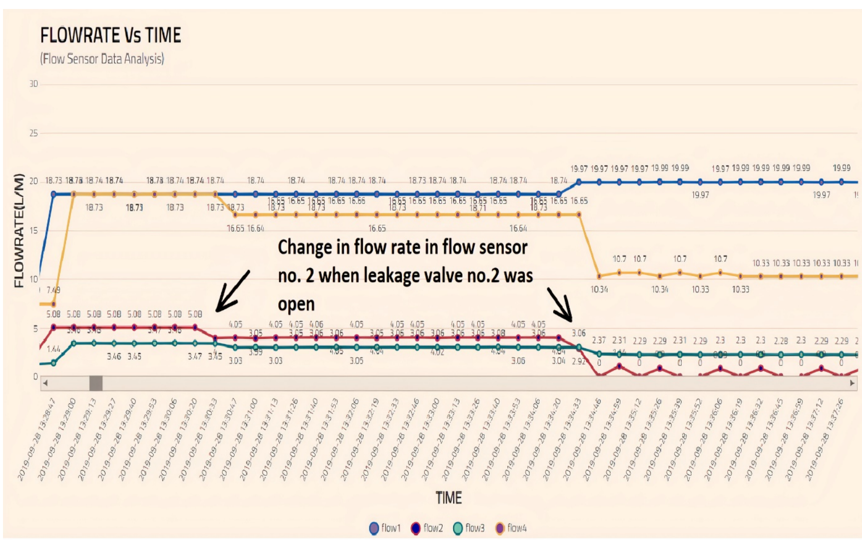

Fig 4. Flow sensor graph when leakage valve no. 2 was open in loop 1

In the most of the research work, simulated data was used which may not give a true picture of the real scenario while some research works were based on real field case which has a fixed type of database because variations can't be possible in the field networks of pipe. Keeping in view of above limitations, if any proven solution based on experimental data is available to locate the leakage, that will provide a great relief to various stake holders associated with distribution work of water through network of pipe. In this research, we attempted to give a unique solution with some limitations to implement in the majority of cases of well-defined loop networks for the water distribution of smart cities. Real-time data was considered from different experiments with many variations made in the laboratory model. To locate and detect leakage, effect of two variables flow and pressure cumulatively was taken into considerations. Total Six numbers of wireless pressure sensors and four numbers of wireless flow sensors were used in a combined manner for the collection of the real-time data from each and every pipe used in the model. All

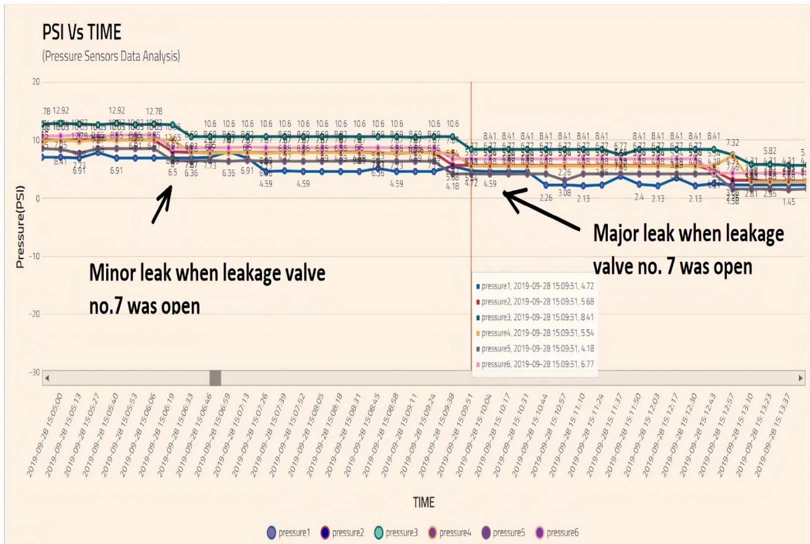

Fig 5. Pressure sensor graph when leakage valve no. 7 was open in loop 2

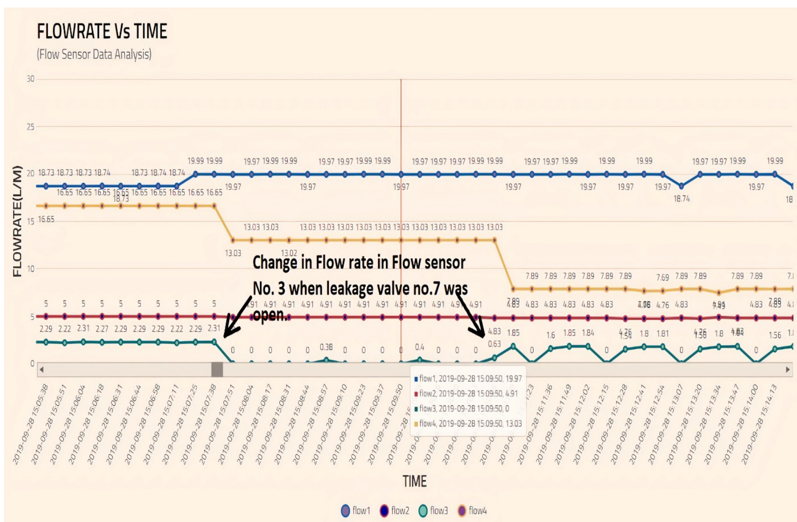

Fig 6. Flow sensor graph when leakage valve no. 7 was open in loop 2

the sensors were deployed very near to the leakage positions. Different variations made in experimental model to overcome the limitations of real field cases as well as simulation data by different software for the results. Different leak size ranging from 0% to 90% of pipe diameter were considered for better suitability in small and large leak cases of the field. In other research works, a linear pipe or may be one branching of pipe was taken for experiments while in this research a well-defined loop network of pipe was considered as observed in the majority of smart cities. Variations of pressure ranging from 0 PSI 30 PSI and flow rate from 0 to 25 LPM were considered in experiments rather than fixed value considered in simulation data. Not much of the research work has been carried out with the consideration of profile level of pipe and cross-section of pipes but both the changes are included in this research. We used leveled profile, gentle slope profile as well as steep slope profile in both uniform and non-uniform cross-section models of experiments. Data generated from all possible different variations and they were well mixed up for testing and training of machine algorithm in this study. Total data were divided into two parts and 90% of it was used for training and 10% was kept for testing. The developed system is also able to find even a small leak immediately as it is presented in pressure and flow graphs as shown in figures. we achieved an average accuracy of 78% to locate the leakage with the use of K-fold approach. Construction of GUI done in such a way that any medium-skill persons operate it easily to locate leakage based on data received from wireless sensors.

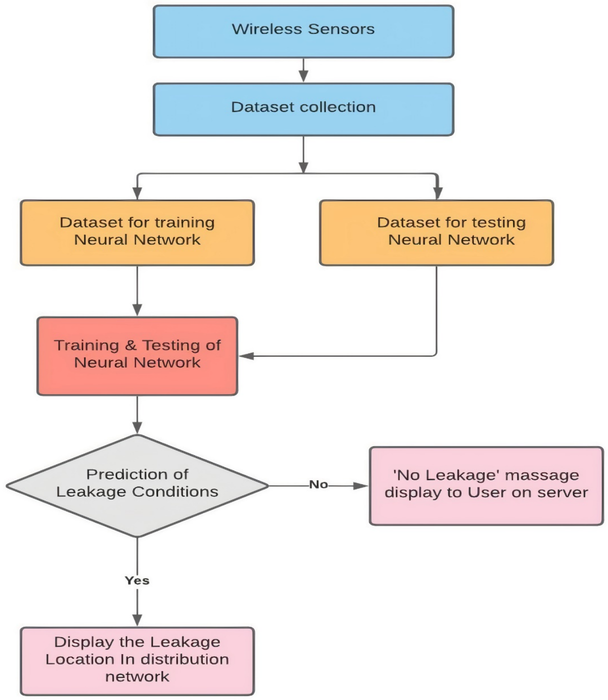

Fig 7. Flow diagram for prediction of leakage.

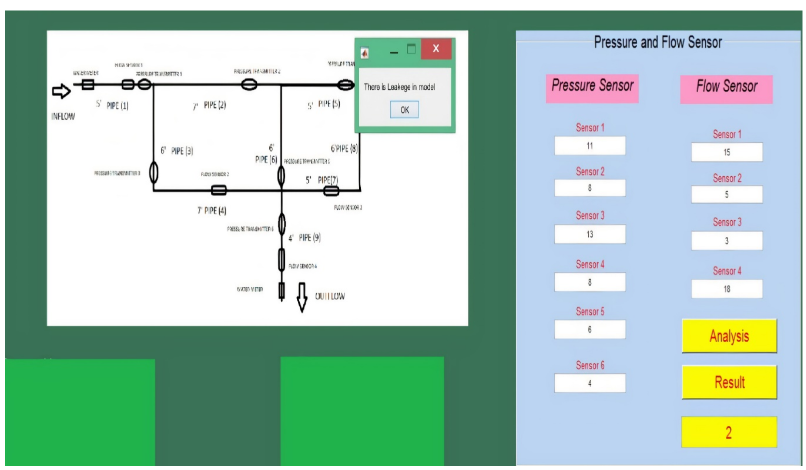

Fig 8. Graphical user interface for prediction of leakage.

4 Conclusion

In this research, Wireless Sensor Network with the use of K-fold machine learning algorithm was used to predict the location of leakage based on several variations made in the experimental model. The selection of the WSN method is due to its effectiveness in real-time data collection and the choice of the K-fold approach was made for algorithm looking towards its unbiased and realistic estimation from the huge collected data. The gross efficiency from this research to locate leakage is achieved as 78% from different variations made in the model that is the size of the leak, pressure variation, profile of pipe network, uniform and non-uniform c/s of pipes. An effort has been made to locate leakage with the use of machine learning algorithms by experiments as it is very difficult to implement in real network of pipes. The novelty of the research is its suitability in different cases based on real-time data of experiments as well as average accuracy was presented rather than taking a single variation in most of the research already done. Beyond that in the model eight different leakage positions were created and six pressure sensors and four flow sensors were used to record the changes in flow parameters of each pipe of the network. The current experiments have some limitations such as water level in pipe considered as full because this is generally observed in the pressure distribution system. Loop network for the distribution of pipe has been considered looking that every smart and big city of the world have well-organized loop piping network. Due to the pandemic situation, a field test has not been performed but the same values of pressure and flow were maintained in experiments as received from Ahmedabad Municipal Corporation (AMC) for distribution of water in Ahmedabad city. This research work can be taken further with implementation in a real case and maybe the use of other alternative machine learning algorithms to achieve higher efficiency as compared to obtaining efficiency in this research.

References

1) El-Zahab S, Zayed T. Leak detection in water distribution networks: an introductory overview. *Smart Water*. 2019;4(1).
2) Kammoun M, Kammoun A, Abid M. Experiments based comparative evaluations of machine learning techniques for leak detection in water distribution systems. *Water Supply*. 2022;22(1):628–642. Available from: https://doi.org/10.2166/ws.2021.248.
3) Lee CWW, Yoo DGG. Development of Leakage Detection Model and Its Application for Water Distribution Networks Using RNN-LSTM. *Sustainability*. 2021;13(16):9262–9262. Available from: https://doi.org/10.3390/su13169262.
4) Mashhadi N, Shahrour I, Attoue N, Khattabi JE, Aljer A. Use of Machine Learning for Leak Detection and Localization in Water Distribution Systems. *Smart Cities*. 2021;4(4):1293–1315. Available from: https://doi.org/10.3390/smartcities4040069.
5) Ravichandran T, Gavahi K, Ponnambalam K, Burtea V, Mousavi SJ. Ensemble-based machine learning approach for improved leak detection in water mains. *Journal of Hydroinformatics*. 2021;23(2):307–323. Available from: https://doi.org/10.2166/HYDRO.2021.093.
6) Idachaba F, Rabiei M. Current technologies and the applications of data analytics for crude oil leak detection in surface pipelines. *Journal of Pipeline Science and Engineering*. 2021;1(4):436–451. Available from: https://doi.org/10.1016/j.jpse.2021.10.001.
7) Liu Y, Ma X, Li Y, Tie Y, Zhang Y, Gao J. Water Pipeline Leakage Detection Based on Machine Learning and Wireless Sensor Networks. *Sensors*. 2019;19(23):5086–5086. Available from: https://doi.org/10.3390/s19235086.
8) Liu M, Yang J, Zheng W. Leak Detection in Water Pipes Based on Maximum Entropy Version of Least Square Twin K-Class Support Vector Machine. *Entropy*;23(10):1247–1247. doi:10.3390/e23101247.
9) Zhou M, Yang Y, Xu Y, Hu Y, Cai Y, Lin J, et al. A Pipeline Leak Detection and Localization Approach Based on Ensemble TL1DCNN. *IEEE Access*. 2021;9:47565–47578. doi:10.1109/ACCESS.2021.3068292.
10) STM International, "Standard Specification for Poly(Vinyl Chloride) Plastic Pipe. 2012. Available from: https://doi.org/10.1520/D1785-12.
11) Liu Y, Ma X, Li Y, Tie Y, Zhang Y, Gao J. Water Pipeline Leakage Detection Based on Machine Learning and Wireless Sensor Networks. *Sensors*. 2019;19(23):5086–5086. doi:10.3390/s19235086.
12) Coelho J, Glória A, Sebastião P. Precise Water Leak Detection Using Machine Learning and Real-Time Sensor Data. *IoT*. 2020;1(2). Available from: https://doi.org/10.3390/iot1020026.