Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7764682

WATER LEAKAGE DETECTION SYSTEM USING MACHINE LEARNING ALGORITHM

Naga Bhargavi Lakshmi Narasu Praharaju
PG Research Scholar, Department of Computer Science Engineering, Raghu Engineering College, Dakamarri, Visakhapatnam – 531162, Andhra Pradesh, India.
vnagabhargavi@gmail.com

Anil Kumar Prathipati
Assistant Professor, Department of Computer Science Engineering, Raghu Engineering College, Dakamarri, Visakhapatnam – 531162, Andhra Pradesh, India.
anilkumarprathipati@gmail.com

Nalinikanth Vemulakonda
vnalinikanth@gmail.com

ABSTRACT:
Water, an essential resource for human survival, must be secured and efficiently used in concert with sustainable development to give a future generation without scarcity. Even though the usage is done efficiently, the stage in which the water reaches mankind from the source experiences some leakages in the system causing the water to get wasted even before reaching the destination. The optimum use of developed technologies and human efforts can cause a decline in the wastage of water. The manual intervention of detecting the leakage is done by analysing the sound of the water leakage area through various devices and replacing the parts with suitable parts. But this traditional way of detecting the leakage causes manual assistance and is comprised of heavy workloads causing high time consumption. Instead, the detection can be done automatically with minimal human assistance reducing time and manual workload. This paper provides a solution to one such extent by enabling the automatic detection of water leakage via suitable microphones as sensor input. Then processing the sound signal for converting the data from the time domain to the frequency domain. The processed data is given to the Machine Learning (ML) model for identifying the water leakage. Three different ML models including KNN (K Nearest Neighbour), SVM (Support Vector Machine), and Random Forest (RF) are employed. Among the three ML models, the best one is identified using the metrics. Finally, the best model is deployed in the mobile app which makes the water leakage detection process simple.
KEYWORDS: Water leakage, Sensor, Machine Learning, Mobile Application, Accuracy

INTRODUCTION:

Water is the most essential component on earth for human beings as well as other living organisms. Other living organisms use water as a primary source for their livelihood whereas humans, one step ahead use water as the source of development and technology. Since industrialization and urbanization, the water resource is controlled by mankind and optimum use has been implemented in developing areas. This moved the water effect according to the needs of the usage of mankind which made the transportation of water sources employing canals and integrated pipelines for efficient employment. Not only for industrialization but also household connections and residential purposes, pipelines are implemented to reach out to every nook and corner. This makes the water supply one of the economic assets of any region. The pipelines are made in a vision that any external force or disturbances should not affect or damage the system at any cost since the replacement of these lines may include difficulties. Most of the pipelines are transported and implemented under the ground to keep them away from any damage that leads to any effect. Since these pipelines are underground, any impact on the lines such as truck transportation on the way of the pipes, natural calamity effects such as soil erosion, floods, earthquakes, building collapse, and others can damage the system and the leakage begins which in turn affects the water transportation. In such cases, the effects must be controlled. Natural calamities are unpredictable in these circumstances. So, the leakage is the point at which the remedy should be identified.

The main causes of water leaks are the degradation and aging of water pipes. Sinkholes could develop owing to ground loss if such leaks are not corrected for an extended period. The pipes withstand the physical and chemical effects to a certain level which might reach its threshold during the aging period. Leaks should therefore be found and rectified as soon as possible. However, because water pipes are underground, leaks cannot be found until obvious ground damage are present. While the pace of renewing the pipelines has been progressively declining, the proportion of pipes that need to be replaced due to age is rising every year [1]. Over 40% of the water in the supply system is lost in some nations owing to water leaks in the supply network. Many nations consider reducing water leaks to be a top priority since doing so will lower the cost and energy needed to produce and pump water, as well as enhance system reliability and satisfy customer demands [2].

In the case of water networks, the development of leaks may allow the admission of viruses and toxins from the environment, potentially endangering human life. As underground cables are also available, the leakage may seep through the soil and may react with the cables causing negative consequences. Shutting off the water system and using acoustic instruments to see if the sound can travel to the end of the pipe without losing strength indicates that the pipes are leak-free and is the traditional method for finding leaks in pipes. These kits, which are offered for sale and are frequently used by plumbers and inspectors, are portable equipment that can hear pipe defects. The enhancement of the irrigated field might suffer because of this solution, which not only necessitates more labour but also implies a pause in the system's regular operations. This type of analysis is therefore no longer real-time but rather a periodic check-up.

Another approach that is frequently utilized is a visual investigation, not only of the pipes when they are visible above ground but also of any places that show indications of flooding of a burst pipe beneath. Like the previous approach, this requires more personnel and cannot be done in real-time. The manual intervention should be done all over the system which may take a huge workload on the workers as well. Hence this paper reveals an idea that can act upon the detection of the leakage without human assistance and intervention, reducing the time consumed and simplifying the workload. The way of implementing the idea can make the detection automatic, such that the process gets simplified, and the replacement outcome of the damaged pipes can be increased at faster rates. As soon as the detection is performed it must be communicated to the respective authorities to work on the issue and to provide remedial and replacement measures. This can be achieved by sending the information through IoT services and enabling efficient and faster modes of communication thus reducing personnel assistance and emerging as the time-saving way.

LITERATURE REVIEW:

The article [3] employs a system that uses temperature change to identify tiny fluid leakage in a PIP system is created. By integrating a fibre optic element on the pipe, the Distributed temperature sensing system calculates the temperature data at multiple points. In addition to that, utilizing Distributed temperature sensing system data and ML analysis, a system was able to check a little fluid leakage even when the temperature of the working fluid changes. The created Pipe in pipe leakage detection technology is deployable in industrial safety-critical systems for self-leakage detection. The article [4] offers a decision-support tool for controlling indoor water leaks. The suggested modelling framework covers data pre-processing and integration, forecasting, interpreting, and spatial mapping and is relevant to a wide range of urban issues that call for an in-depth examination of their spatial properties. The study [5] describes the creation of a cloud-based information management system for an AI-based water leak-detecting system. The system may automatically collect and compare leakage sounds and produce a model that is accessed by a mobile app. The design and implementation of a leaking sound gathering were successfully managed to be implemented. Datasets for leakage sounds were gathered from the Metropolitan Waterworks Authority in several different locations. DNN, CNN, and SVM were developed and compared. The challenge in leak identification is presented in the paper [6] as a potential use of geometric deep-learning techniques. To estimate nodal pressures in a WDN, it is proposed to train Chebyshev Convolutional Networks. An O/P signal is acquired to find leaks by comparing the two network outputs. Based on IoT technology, the article [7] serves as a water management concept. The development of a Water Distribution Network (WDN) abstraction prototype. On the network, sensors are put in place to record the desired physical values, such as flow rates, turbidity, and pH levels of the water. after the creation of the sensor network, transmit the data to the Firebase platform. Finally, a full IoT testbed architecture is suggested to connect all the IoT components.

The suggested method in the paper [8] accurately anticipated the corrosion severity levels by combining the detection of corrosion through acoustic signals from accelerated testing with ML techniques. As a result, detection is much easier than using conventional approaches since the leakage area is more precisely defined. The study [9] objective was to demonstrate the feasibility of detecting water leaks in underground tunnels by remote sensing. Field experiments showed that using the intensity and spatial details of tunnel point cloud data, it is possible to quickly collect, find filters, and record locations of water leakage. This is especially helpful because concrete degradation from water infiltration is known to occur early and cause corrosion, spalling, and strength loss. The LiDAR scanner employed here demonstrated its ability to shorten survey times. The wave propagation model serves as the foundation for the explainable ensemble tree model presented in the research [10], which is based on an optimized feature space. The piecewise spectrum entropy is proposed and utilized to build the feature space after the vibration signal has been processed. Four ensemble tree models are then utilized to create leakage identification models after the Boruta method is used to reduce the number of features. The progress of ML applications in both natural and manmade water systems is thoroughly summarised in the paper [11]. According to their structures and mechanisms, the benefits and drawbacks of frequently used algorithms are examined, and a suitable ML algorithm is suggested regarding various studies as well as future directions for the technologies and advancement of ML in the field of the water system. The study [12] describes a way for utilizing ANN to find and pinpoint pipeline water leaks. The leak position is calculated using the proposed method by estimating the pipe's friction factor as an input. The data-training set for the ANN was enhanced using data produced by a simulator. In a prototype plant, the algorithm underwent experimental testing. The outcomes show that the suggested strategy is practical and performs well.

METHODOLOGY:

The traditional way of detecting water leakage using the manual workload is time-consuming and takes a lot of manpower. This method is done by analysing every nook and corner of the pipeline system to check the fullest of the pipeline system. This can be difficult in some cases as the leakage may have occurred in the areas where the manual intervention cannot be done such as tunnels, railway lines, dams to domestic purpose areas, underground, etc. Hence the proposed solution can be implemented in those areas which makes it much easier than the traditional methods of leakage detection. The comparison of the traditional and proposed method is depicted in figure 1.

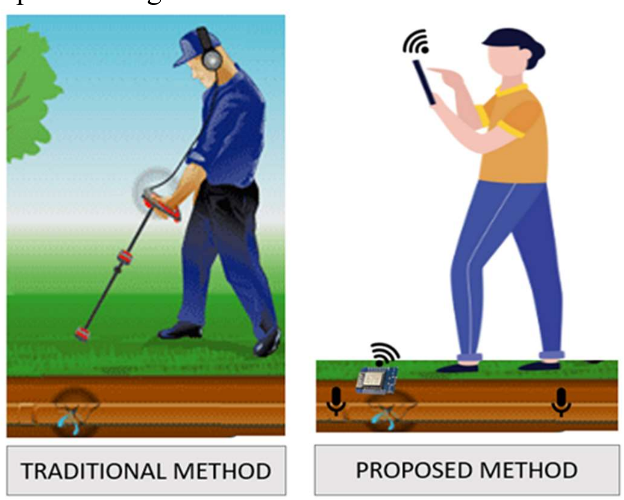

Fig 1: Traditional method vs proposed method.

As soon as the pipeline is laid in the areas, the sensors which act as the input sources can be employed in those areas as well making it available every time to analyse the sound released from the pipeline when the leakage is detected. The collected sound signal is in the time domain, and it is very difficult to process further. So, the time domain data is converted to the frequency domain using the transform technique. Next, the three various ML models are developed, trained, and tested. The outcome of the ML model in the test phase is compared to identify the best one. Finally, the best ML model is deployed in the mobile app. The methodology of the proposed solution is described in figure 2.

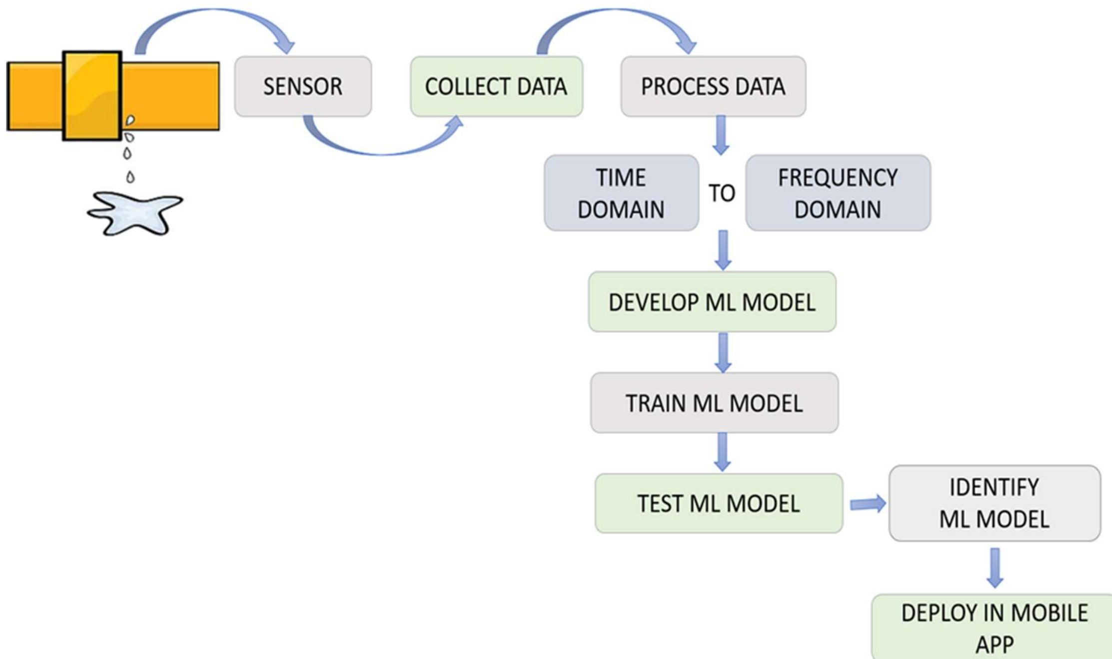

Fig 2: The working steps of the proposed solution

A. Data Collection and Processing

Water leak detection can be done using the sensor device (microphone) as the input which sends the data to the cloud database where the data gets stored. The sensors are employed at specific intervals of distance according to the accuracy and working capacity of the sensor. When the leakage happens, the sensor reads the input from the noise, then the audio signals are converted into numerical data.

The data is then collected and processed by a mobile application which is developed according to the needs of the proposed solution. The process data will be available in the time domain format which is the basic acquired unprocessed data from the sensor to the cloud which is collected at specific intervals of time. The data in turn is converted in the form of frequency domain signals. The conversion is done by the Fast Fourier Transform (FFT) method to make the signal loss-free and accurate. This conversion is due to the feasibility and reality of the data in the frequency domain as compared to the time domain. The received data is then inserted into the developed ML models such as KNN, SVM, and RF.

The data is collected in two different conditions leakage and non-leakage. These data are used to train the ML model. The leakage and the non-leakage data are given in figure 3 and will be used as the input for the models.

	A	B
1	Time	Non-leakage data
2	18-03-2022 14:03:33.522	49.885746
3	18-03-2022 14:03:33.522	49.885746
4	18-03-2022 14:03:33.523	49.885746
5	18-03-2022 14:03:33.523	49.885746
6	18-03-2022 14:03:33.524	49.885746
7	18-03-2022 14:03:33.525	49.885746
8	18-03-2022 14:03:33.525	49.885746
9	18-03-2022 14:03:33.526	49.885746
10	18-03-2022 14:03:33.526	49.885746
11	18-03-2022 14:03:33.527	49.885746
12	18-03-2022 14:03:33.528	49.885746
13	18-03-2022 14:03:33.528	49.885746
14	18-03-2022 14:03:33.529	49.885746
15	18-03-2022 14:03:33.529	49.885746
16	18-03-2022 14:03:33.530	49.885746
17	18-03-2022 14:03:33.531	49.885746
18	18-03-2022 14:03:33.531	49.885746
19	18-03-2022 14:03:33.532	49.885746
20	18-03-2022 14:03:33.532	49.885746

	A	B
1	Time	Leakage data
2	11-03-2022 15:49:40.734	0.001897
3	11-03-2022 15:49:40.734	0.003813
4	11-03-2022 15:49:40.735	-0.001362
5	11-03-2022 15:49:40.735	0.003247
6	11-03-2022 15:49:40.736	0.001409
7	11-03-2022 15:49:40.737	0.000553
8	11-03-2022 15:49:40.737	-0.000333
9	11-03-2022 15:49:40.738	0.002415
10	11-03-2022 15:49:40.738	0.005157
11	11-03-2022 15:49:40.739	0.00618
12	11-03-2022 15:49:40.740	0.003705
13	11-03-2022 15:49:40.740	0.008518
14	11-03-2022 15:49:40.741	-0.004402
15	11-03-2022 15:49:40.742	0.003224
16	11-03-2022 15:49:40.742	-0.00348
17	11-03-2022 15:49:40.743	-0.002272
18	11-03-2022 15:49:40.743	-0.001957
19	11-03-2022 15:49:40.744	-0.006157
20	11-03-2022 15:49:40.745	-0.004717

Fig 3: Collected data: Leakage data vs non-leakage data.

The collected data set is separated in the ratio of 8:2 and made as an input to train the ML model. A total of around 1500 samples were collected for each of the leakage and non-leakage process. The trained data takes some time to complete the training process. Then the remaining data is made as input which employs in the testing of the ML model. It is suggested to use the training model in higher composition than the testing data, to improve the accuracy of the ML model. By comparing the outcome of ML models, it will be helpful to identify which one will be much more accurate than others. The best ML model is then deployed in the mobile application. The mobile application indicates the water leakage as well as the area nearer to the sensor where the leakage has happened.

MACHINE LEARNING MODEL:

The ML models such as KNN, SVM, and RF are used in the proposal to find the leakage detection values.

A. KNN

KNN in ML is one of the essential algorithms that comes to supervised learning techniques. It is mainly used in fields of data mining, pattern identification, etc. The algorithm is mainly used when the prediction of the values is to be done in a specific. They are also induced with the capacity of working with large numbers of data sets thus making it one of the robust algorithms in the supervised learning areas [13]. Through a process of "feature similarity" or "nearest neighbours," the K-nearest neighbours (KNN) method can determine which group a new data point would belong to. When great accuracy is required and there are a few unknown points to categorize, KNN is quite helpful. KNN does not use the learning principle. Because other models, require time for training, KNN is the preferred approach when we need a quick prediction for a small set of data points.

B. SVM

Support Vector Machine, or SVM, is among the most popular supervised algorithms designed to directly address issues of classification and regression. Most of its applications, however, are in the realm of ML Classification [14]. To efficiently categorize new data points in the upcoming times, the SVM method seeks to find the optimum line that really can partition an n-dimensional space into categories. This optimal decision boundary is referred to as a hyperplane. SVMs can handle high-dimensional data and do well with tiny datasets, which is one of their benefits.

C. RANDOM FOREST

The results from multiple decision trees are combined into one using the famous ML technique known as RF. RF is commonly used in data science competitions and the real world. Its flexibility and ease of use in resolving classification and regression problems have contributed to its broad adoption [15]. They frequently have high accuracy, don't need categorical feature encoding, or feature scaling, and only require little parameter tuning.

MOBILE APPLICATION DEVELOPMENT

The data will be displayed as an output to the mobile application which will be much more feasible, accessible, and reliable. The cloud data which is received from the sensors are accessed by the firebase system which in turn works for the display medium. It is a type of app development region in which the developers can access the specifications which can make the data much more reliable. It can be used to build products, release, and monitor the products and engage with the access. Various services and real-time database applications are available in the firebase network to make the system much more user-friendly.

Because App Inventor is a cloud-based program, you can make mobile or tablet applications directly from your web browser. The assistance provided by the MIT app inventor is Setup Guidelines for Testing, an Overview of the Designer, Blocks Editor, Packaging, and Sharing Apps. Several well-known Android phone models as well as the Mac OS X, GNU/Linux, and Windows operating systems are supported by the App Inventor development environment. Any Android phone can run applications downloaded via App Inventor. Using the above facilities, users can customize the app as per the needs required to make it feasible, reliable, accessible, and user-friendly.

RESULT AND DISCUSSION

The water leakage, as well as non-leakage data, are collected by placing the sensors in the pipeline. The collected data is composed of equal distribution of water leakage and non-leakage samples. Those data are processed and split into two parts (80% for train and 20% for testing). After training the ML model by giving the processed data, the ML model will be tested. The outcome of the ML model in the testing phase is compared with the actual outcome. The metrics like accuracy, True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR), precision, and F1-score are chosen to assess the ML model. The metrics score of all three models is given in table 1.

Table 1: Comparison of values of each algorithm

MODEL	ACCURACY	TNR	TPR	PRECISION	FI-SCORE
KNN	96	97.3154	94.702	97.27891156	95.973154
RF	97.6667	96.6887	98.6577	96.71052632	97.674419

SVM	98.6667	99.3243	98.0263	99.33333333	98.675497

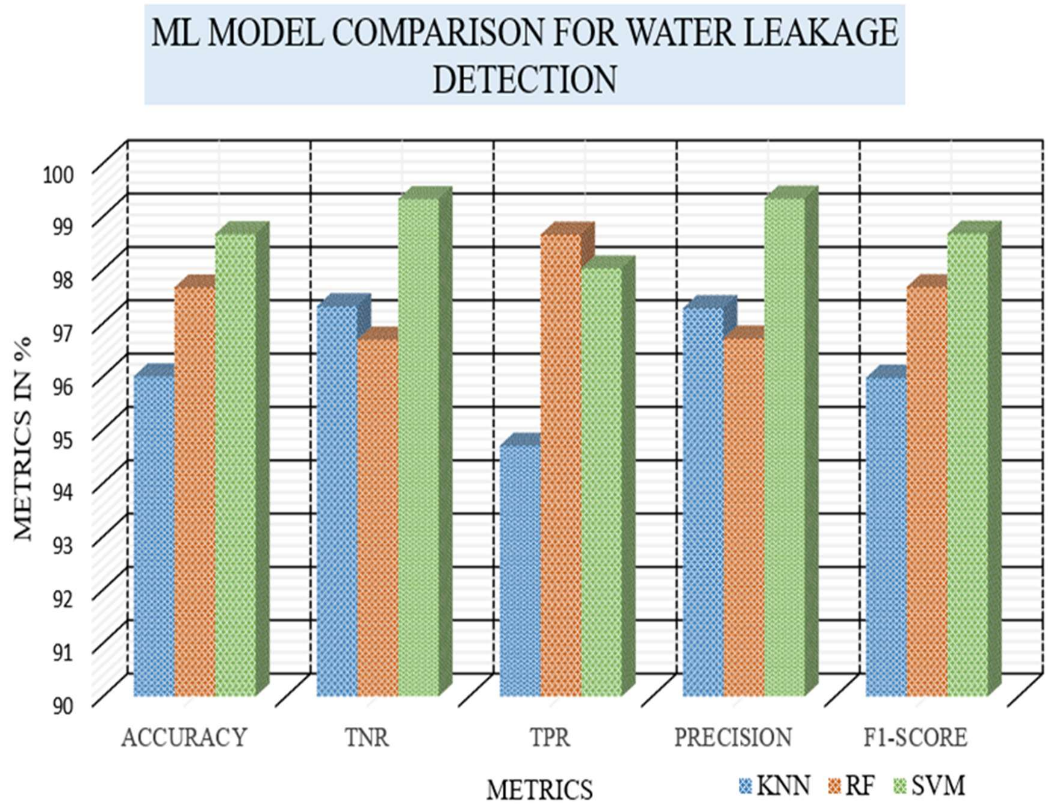

Fig 4: The comparison of the three models in graph form

Figure 4 depicts the comparison of the above-mentioned metrics values attained by three ML algorithms such as KNN, SVM, and RF. With the support of these metrics, the best model can be selected and deployed in the mobile app. The blue bar in the figure indicates the metrics score of KNN, similarly, the red bar indicates RF, and the green bar indicates SVM. From the given table and the graph, it is clearly shown that the SVM algorithm is much better when compared with the other two ML algorithms. The SVM model is deployed in the mobile app for automatic water leakage identification.

The working of the mobile application is represented in figure 5 which is the outlook of the user interface display. This makes the user to identify the leakage and non-leakage areas with the help of sensor location. When the leakage area is detected by the sensor, the location is converted to red color in the mobile app whereas the non-leakage area is depicted in blue color. When the red color is clicked, another interface is opened which directs to the google maps of the mobile. This helps the user to identify the region where the leakage is happening.

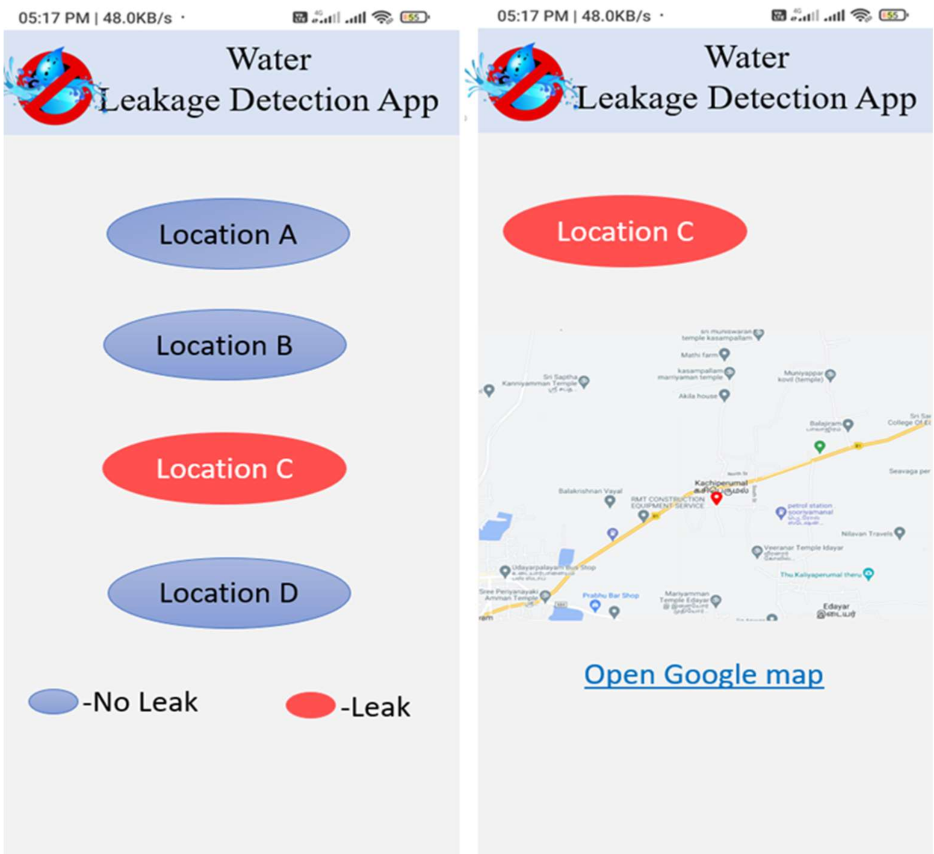

Fig 5: Mobile app development

CONCLUSION:

The water delivery system is among the most vital services for municipalities. Unfortunately, if the underground pipe network deteriorates, a considerable volume of water is wasted each year, usually unreported. Forecasting water pipe leakage remains a difficult subject because of the complex subsurface environment. Most organizations practice repairing pipes when leaks are explicitly recognized, whereas many little leaks go unnoticed till the losses manifest in the form of ground cavitation. Water leakage can result in electric shock through touch, stagnant water-producing infections, and so on. As a result, humans must provide a solution for leakage and reduce the amount of unusable water draining into sewage systems. This study attempted to reduce water leakage by automating the process with an ML approach. To detect leakage, three approaches are used: SVM, KNN, and RF. The SVM model was found to be the best of the three, with the highest accuracy of 98.66%. The SVM model is used in the mobile app to detect the location of a water leak. This research assists society in effectively conserving water.

REFERENCE:

1. Nam, Y. W., Y. Arai, T. Kunizane, and A. Koizumi. "Water leak detection based on the convolutional neural network using actual leak sounds and the hold-out method." Water Supply 21, no. 7 (2021): 3477-3485.

2. Rogers, D. "Leaking water networks: an economic and environmental disaster." Procedia Engineering 70 (2014): 1421-1429.

3. Kim, Hayeol, Jewhan Lee, Taekyeong Kim, Seong Jin Park, and Hyungmo Kim. "Advanced thermal fluid leakage detection system with machine learning algorithm for pipe-in-pipe structure." Case Studies in Thermal Engineering (2023): 102747.

4. Shin, Jihoon, SangHyun Son, and YoonKyung Cha. "Spatial distribution modeling of customer complaints using machine learning for indoor water leakage management." Sustainable Cities and Society 87 (2022): 104255.

5. Vanijjirattikhan, Rangsarit, Sunisa Khomsay, Nathavuth Kitbutrawat, Kittipong Khomsay, Unpong Supakchukul, Sasiya Udomsuk, Jittiwut Suwatthikul, Nutthaphan Oumtrakul, and Kanchanapun Anusart. "AI-based acoustic leak detection in water distribution systems." Results in Engineering 15 (2022): 100557.

6. Garðarsson, Garðar Örn, Francesca Boem, and Laura Toni. "Graph-Based Learning for Leak Detection and Localisation in Water Distribution Networks." IFAC-PapersOnLine 55, no. 6 (2022): 661-666.

7. Ali, Ahmed S., Mahmoud N. Abdelmoez, Mahmoud Heshmat, and Khalil Ibrahim. "A solution for water management and leakage detection problems using IoTs based approach." Internet of Things 18 (2022): 100504.

8. Sheikh, Muhammad Fahad, Khurram Kamal, Faheem Rafique, Salman Sabir, Hassan Zaheer, and Kashif Khan. "Corrosion detection and severity level prediction using acoustic emission and machine learning based approach." Ain Shams Engineering Journal 12, no. 4 (2021): 3891-3903.

9. Hawley, C. J., and P. J. Gräbe. "Water leakage mapping in concrete railway tunnels using LiDAR generated point clouds." Construction and Building Materials 361 (2022): 129644.

10. Xu, Weinan, Shidong Fan, Chunping Wang, Jie Wu, Yunan Yao, and JunChen Wu. "Leakage identification in water pipes using explainable ensemble tree model of vibration signals." Measurement 194 (2022): 110996.

11. Huang, Ruixing, Chengxue Ma, Jun Ma, Xiaoliu Huangfu, and Qiang He. "Machine learning in natural and engineered water systems." Water Research 205 (2021): 117666.

12. Pérez-Pérez, Esvan de Jesús, Francisco Ronay López-Estrada, Guillermo Valencia-Palomo, L. Torres, Vicenç Puig, and Jesus Darío Mina-Antonio. "Leak diagnosis in pipelines using a combined artificial neural network approach." Control Engineering Practice 107 (2021): 104677.

13. Zaidi, Syed Ali Raza. "Nearest neighbor methods and their applications in the design of 5G & beyond wireless networks." ICT Express 7, no. 4 (2021): 414-420.

14. Zhu, Zhenfeng, Xingquan Zhu, Yuefei Guo, Yangdong Ye, and Xiangyang Xue. "Inverse matrix-free incremental proximal support vector machine." Decision support systems 53, no. 3 (2012): 395-405.

15. Cutler, Adele, D. Richard Cutler, and John R. Stevens. "Random forests." Ensemble machine learning: Methods and applications (2012): 157-175.