A REPORT

ON

‘LEAKAGE DETECTION IN SMART WATER DISTRIBUTION SYSTEMS’

BY

Name of the Student	ID Number
Pavni Jaiswal	*2016A8PS0357P*
Neelabh Sinha	*2016B5A80600P*

Prepared on completion of the
Laboratory Project (INSTR F366)

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

December, 2019

ACKNOWLEDGEMENT

First we would like to thank the BITS Pilani administration, including Vice Chancellor Dr. Souvik Bhattacharya and Director Dr. A.K. Sarkar to give us an opportunity to be a part of BITS Pilani, Pilani Campus and to introduce a course as Laboratory Project through which we can get practical industrial exposure.

We would also like to thank CSIR-CEERI, including Dr. Akbar, our mentor Dr. Bhausaheb Ashoke Botre and Senior Research Fellow Ms. Pooja Chaudhary, for giving us a chance to work and gain practical industrial experience. They provided immense support and guidance throughout the project, and their support was the prime reason why this project was successfully implemented.

In the Department of Electrical and Electronics, we would take the opportunity to thank our mentor, Dr. Karunesh Kumar Gupta. The guidance and cooperation shown by him can be second to nothing. Whether it is personal or academic help, he always stood by our side.

Moving along, we would also thank our professors, Dr. Vinod Kumar Chaubey, Dr. Surekha Bhanot, Dr. Puneet Mishra, Dr. Navneet Goyal, who taught us the courses relevant to do this project, which were Neural Network and Fuzzy Logic, Transducers and Measurement, Industrial instrumentation and control, and Machine Learning respectively.

In addition, a big thank you to the staff of the Department of Electrical and Electronics, and officials in CEERI Pilani, particularly those who worked in the smart water distribution system, who supported us in getting along with everything smoothly,m and ensured the smooth functioning of every resource needed.

At last, we would like to thank each other, i.e. the members of the group who came along together to finish this project properly under the given time.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI (RAJASTHAN)

DATE OF SUBMISSION: 29th November, 2019

TITLE OF THE PROJECT: LEAKAGE DETECTION IN SMART WATER DISTRIBUTION SYSTEMS

ID NO.: 2016A8PS0357P **NAME:**PAVNI JAISWAL **DISCIPLINE:** B.E. E&I.

ID NO.: 2016B5A80600P **NAME:** NEELABH SINHA **DISCIPLINE:** M.Sc. Physics + B.E. E&I

Name of Mentor: Dr. B. A. Botre **Designation:** Principal Scientist

Name of Mentor: Dr. K. K. Gupta **Designation:** Associate Professor

KEY WORDS: Smart water grid, Leak detection, Deep neural network, Modelling of hydraulic distribution network.

PROJECT AREAS: Transducers and Measurement Systems, Deep Learning, Industrial Instrumentation and control

ABSTRACT:
This project is the study of various leakage detection techniques, and using an IOT based smart water distribution system for leakage detection. The distribution network is equipped with all relevant transducers, a programmable logic controller, and SCADA system which is used for data acquisition and control. The system is simulated for leakage of various types, and the collected data is used to develop a model based on deep learning techniques to detect leak. A The developed model is then optimized for recall and precision.

TABLE OF CONTENTS

INTRODUCTION

A Smart Water distribution Grid is a two-way real time network with sensors and devices that continuously and remotely monitor the water distribution system. Smart water meters can monitor many different parameters such as pressure, quality, flow rates, temperature, and others.

In this smart model, leakages are unavoidable problems. Many factors can result in the formation of a leak in the system, a few examples can be: aging pipelines, corrosion, excess pressure, operational error, or rapid opening and closing of valves.

Systematic leakage control programmes are a necessity to avoid the extra cost caused in response to a leakage. The cost can be associated with reduction in pipeline pressure, or in extra pumping, treating and transporting of clean water. Leakage control programmes can be performed in two parts, firstly water audits should be done which involves measuring the volume of water moving and identifying the segment which might have a leak. The exact location of the leak can be identified using leak detection surveys.

In present day scenario, various traditional techniques are combined with artificial intelligence to develop robust models which are highly accurate. Through this project, various leakage detection algorithms are studied, and an IoT based water distribution network is set up to study leakage. Further, the traditional techniques are combined with deep learning algorithm to develop a model which successfully detects leak.

The report is divided into seven sections. Section 1 describes various leak detection techniques being used in distribution networks. Section 2 and 3 describes the modelling and experimentation on the smart-water distribution system. Section 4 explains the neural network model used for the training of various parameters used to detect the presence of leak. Section 5 is focused on the observed results and its analysis. Section 6 concludes the analysis. Section 7 mentions the scope of possible improvements in the trained model and the technique of leak detection.

1. TYPES OF LEAK DETECTION TECHNIQUES

Water pipeline leakage detection techniques, can be broadly classified in three categories:

- Conventional techniques
- Software-based techniques
- Hardware-based technique

Conventional leak detection techniques are the traditional method in which an experienced personnel or workers detect and localize leaks. The method includes visual observation, in which any abnormal pattern, noise or smell is detected. Though the method is good for localization of leaks, it's effectiveness depends on the experience of the worker, the size of the leak and the inspection frequency.

Software-based techniques are based on monitoring internal pipeline parameters, such as pressure, flow rate, and temperature. The complexity and reliability of these techniques are very significant. These methods fall behind because the data analysis method is based on pressure, temperature and flow rate from inside the pipeline, and thus, installation of transducers have to be done inside the pipeline system. The methods included in this category, are:

- Hydrostatic testing
- Mass balance
- Pressure point analysis
- Statistical analysis
- Transient analysis
- Multiple model algorithm

Hardware-based techniques are based on appropriate equipment for leak detection and location. They can be further classified in contact or non-contact methods. Hardware systems have high cost of installation and maintenance. The methods included in this category, are:

- Tracer gas injection
- Infra-red thermography
- Ground penetrating radar
- Acoustic leak detection
- Pipe injection gauge based
- Fibre-optics method
- Vibration method
- Bi-wire sensor
- Temperature gradient analysis

The key attributes of these methods can be summarized as:

Table I: Attributes for Conventional and Software methods

Method	Leak Sensitivity	Location Estimate	Operational Change	Availability	False Alarm	Maintenance Requirement	Cost
CONVENTIONAL METHOD	High	No	No	Yes	Medium	Medium	High
SOFTWARE METHODS							
Hydrostatics	High	No	No	Yes	High	High	High
Mass Balance	Low	No	No	Yes	High	Low	Low
Pressure Point Analysis	High	No	No	Yes	High	High	High
Statistical Analysis Model	High	No	No	Yes	Medium	Medium	High
Transient based Method	High	No	No	Yes	Medium	Medium	High
Multiple model algorithm	High	Yes	No	Yes	Medium	Medium	High

Table II: Attributes for Hardware methods

HARDWARE METHODS							
Visual Observation	High	Yes	No	Yes	High	Low	Medium
Tracer gas Injection	High	Yes	No	Yes	Low	High	High
Infra-red thermography	High	Yes	No	Yes	Medium	High	High
Ground penetrating Radar	High	No	No	Yes	High	Medium	High
Acoustics	High	Yes	No	Yes	High	Medium	High
Pipe Inspection Gauge	High	Yes	No	Yes	Medium	Medium	High
Vibration	High	Yes	No	Yes	Medium	Low	Low
Leak detection bi-wire sensors	High	Yes	No	Yes	High	High	Medium

In this project, the multiple model algorithm will be combined with AI techniques for leak detection in the distribution system. A multiple-model algorithm for a system with flow sensors is one of the methods to detect leakage, which analyzes the water inside the pipeline. The implementation of hardware and software in the system follows the latest trend of using the Internet of Things (IoT) to demonstrate the use of such architecture in a real-time process control. The advantage of the system is that the energy efficiency and lifetime of the equipment can be improved by using a multimodel approach. However, installing the flow rate sensor inside the pipeline for measuring the water flow directly is a disadvantage, as it is a hassle for the existing water pipeline system.

2. MODELLING OF HYDRAULIC NETWORKS

A hydraulic distribution network comprises of a collection of interconnected pipe sections, pumps and accessories such as elbows and valves. End point of each pipe section is union node or fixed grade node. A point where two or more sections are joined together is called a joint node. A point where constant piezometric height is maintained is called a node of fixed degree.

The MATLAB Simulink model for the used hydraulic network used for the data collection is:

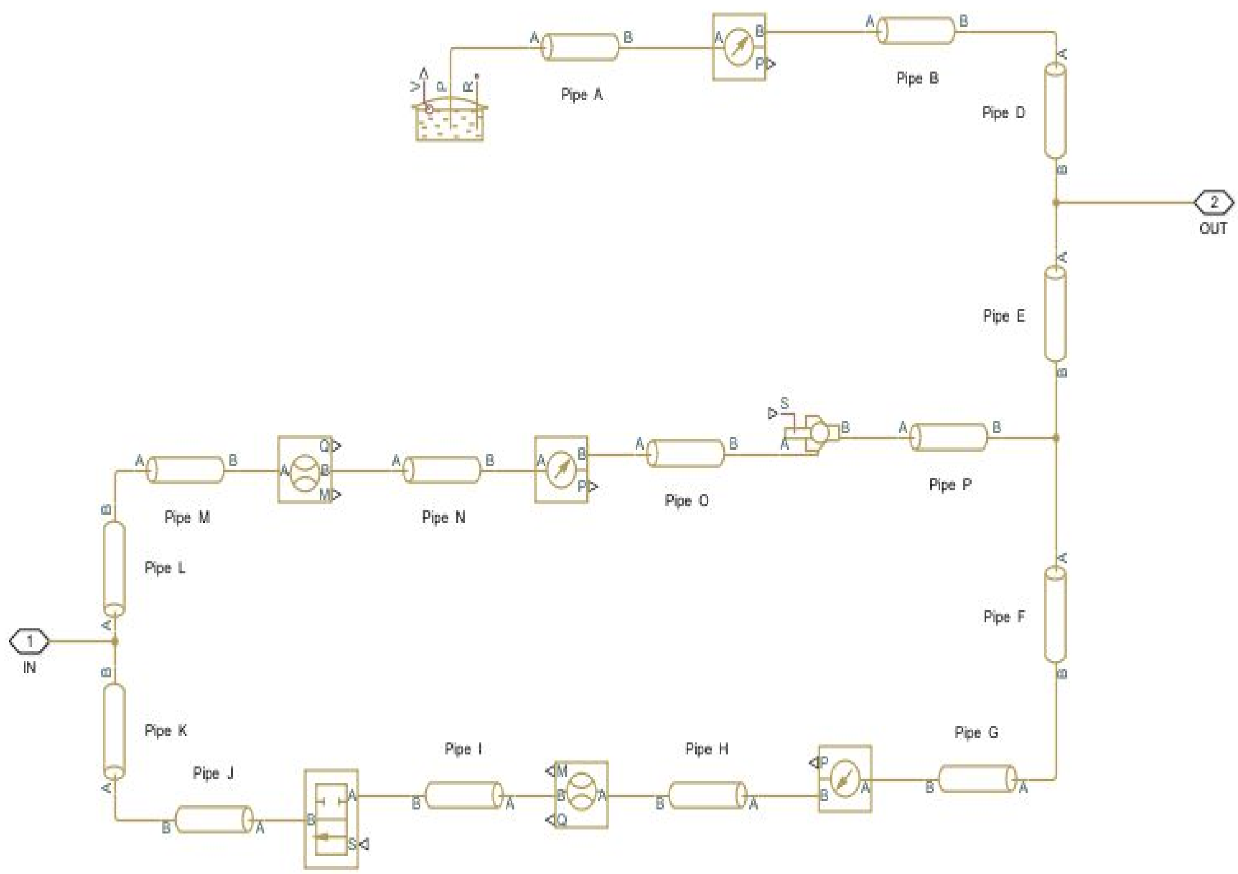

Figure 1: Ground Layer

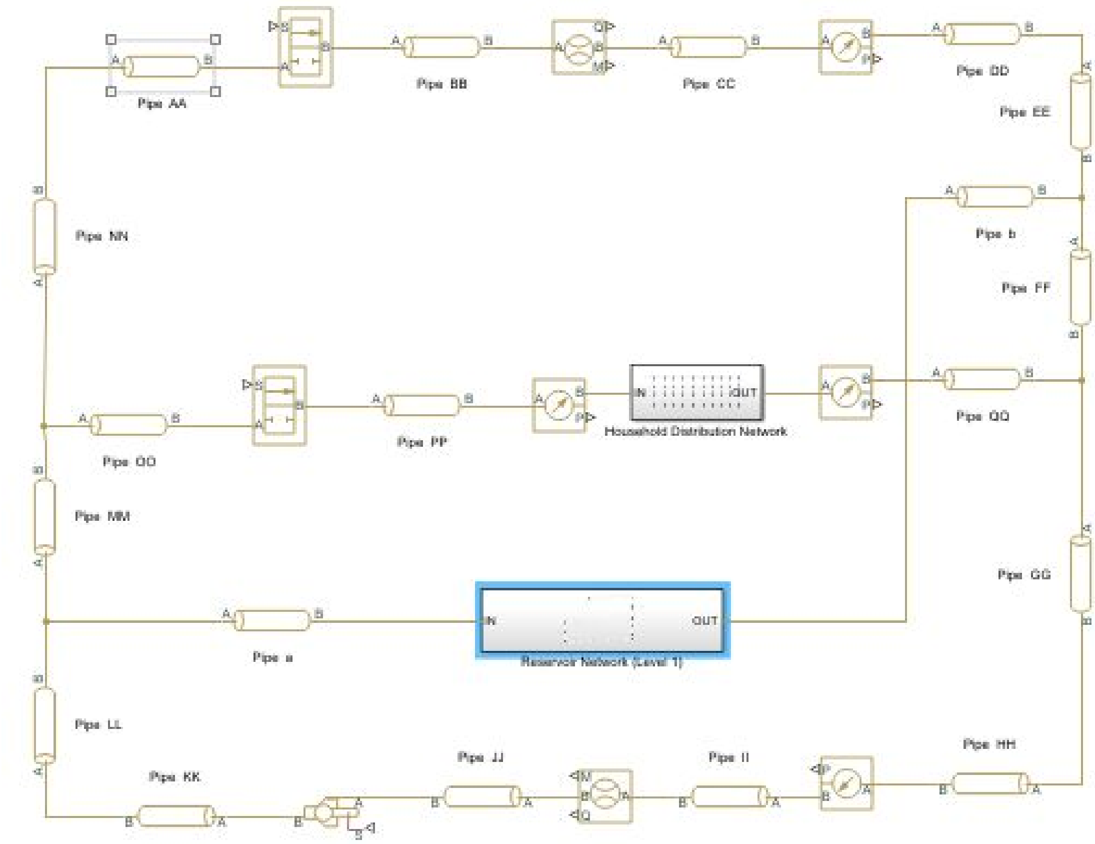

Figure 2: First Layer

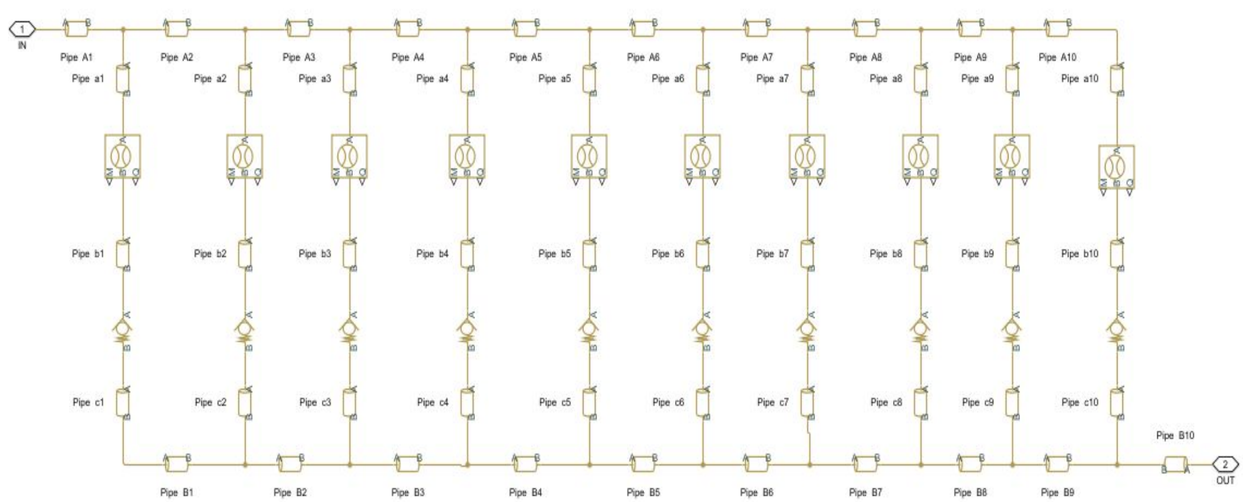

Figure 3: Distribution network in First Layer

The numerical equations associated with the hydraulic model are:

$$N_P = N_J + N_L + N_F - 1 \qquad H_j - H_i = \mu P_u / Q_{ij}$$

$$H_j - H_i = a_0\eta^2 + b_0\eta Q_{ij} + c_0 Q_{ij}^2$$

$$\epsilon = k_s / d$$

$$f = \begin{cases} 64/\mathrm{Re}, & \mathrm{Re} < 2000 \\ \left(-2\log_{10}\left(\frac{\epsilon}{3.7} + \frac{5.74}{\mathrm{Re}^{0.9}}\right)\right)^{-2}, & \mathrm{Re} > 4000 \end{cases}$$

$$f = 0.25\left(\log_{10}\left(\frac{\varepsilon}{3.7} + \frac{5.74}{\mathrm{Re}^{0.9}}\right)\right)^{-2}$$

$$H_i - H_j = h_{ij} = fLQ_{ij}^2 / 2gA^2 d \qquad r_{ij} = fL/\left(2gA^2 d\right)$$

$$Q_f = \lambda_f \sqrt{H_f} \qquad \lambda_f = c_d A_d \sqrt{2g}$$

Where,

NP: No. of pipe sections,
NJ: No. of union nodes,
NL: No. of closed loops,
NF: No. of fixed degree nodes,
Hj: Pressure head,
Pu: Useful power,
Q: Flow,
n: Proportion of rotational speed,
E: Relative roughness,
ks: Absolute roughness,
d: Diameter,
Λf: Leakage coefficient,
cd: Discharge coefficient,
g: Gravity,
A: Area of the pipe.

The model was then simulated and the following inferences were drawn:

- **FLOW**
  The sign of change in flow is opposite on either side of the leak point, which is used to localize the leak. The change is maximum on the pipe segments near the leak, and decreases gradually. The change is minimum in a pipe segment which does not connect leak point and a destination, along the direction of water flow.

- **PRESSURE**
  The pressure decreases at all non-boundary points whenever there is a leak. The change in pressure is maximum at the point of leakage, and decreases as we move away from the leak point. The pressure change is more that point comes before the leak point along the direction of flow of water.

3. EXPERIMENT

The smart-water grid situated at CSIR-CEERI is a two tier model. It is based on the concept of smart cities. The grid can be operated in four modes, namely Raw Water (RW) mode, Treated Water (TW) mode, Reverse Osmosis (RO) Mode and Descaling Mode. A pump, which can be operated between 30 to 50 Hz, is used to drive water from tanks into the system. The ground layer receives the water from one of the two tanks, namely the raw water tank and the treated water tank. After the water flows through various meters and quality indicators, a Reverse Osmosis (RO) system is installed for treating of the water, which is later available in the smart water vending machine and in the treated water tank. The Smart water grid also has the functionality of descaling, which helps clean the pipes and sensors for long term efficiency of the model.

The model as shown in SCADA software is:

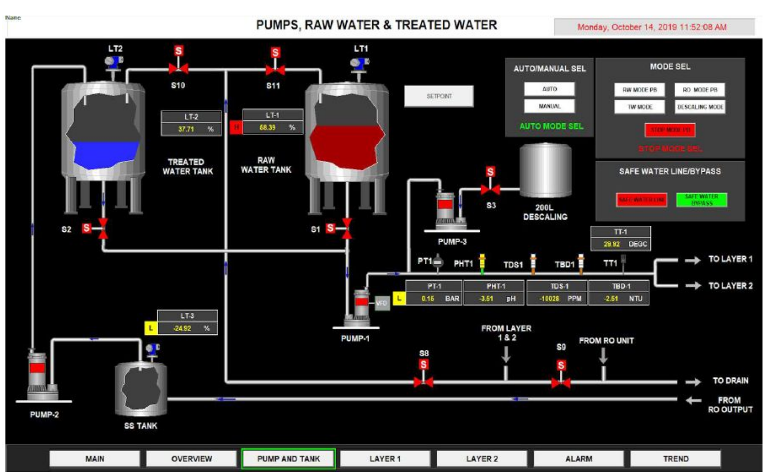

Figure 4: Pumps and Tanks in smart-water grid

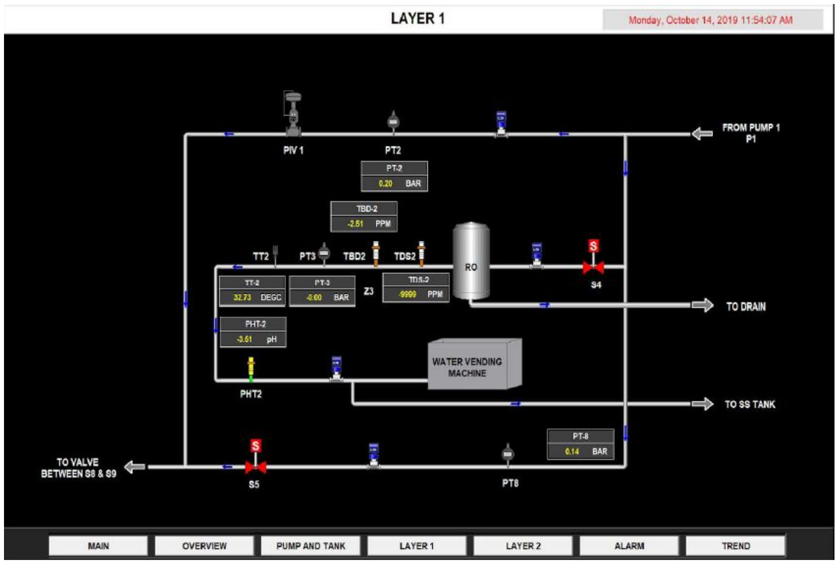

Figure 5: Ground layer of the smart-water grid

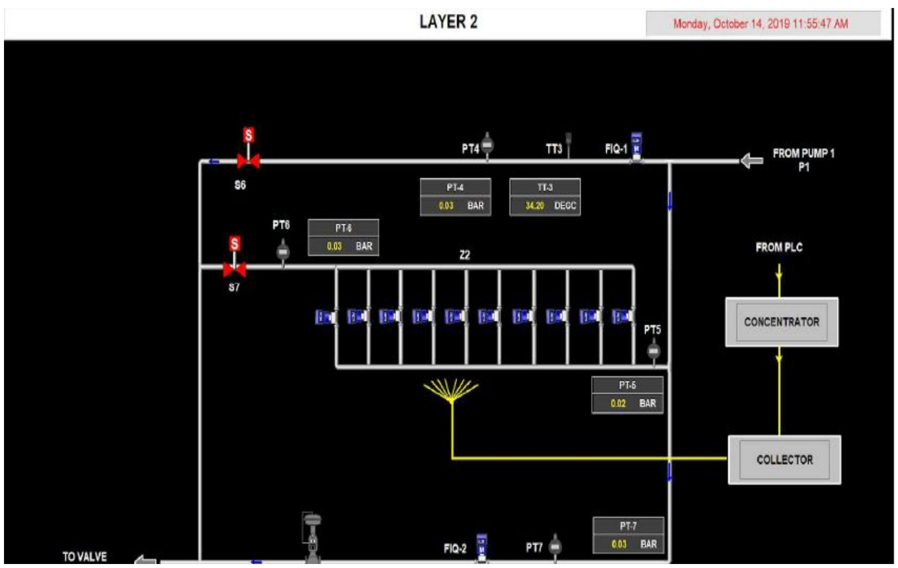

Figure 6: First layer of the smart-water grid

The various sensors and meters in the water grid include Ultrasonic, Electromagnetic and Solar flow meters, pressure and temperature transmitters, pH, TDS and Turbidity indicator, Solenoidal, motorized, PIV and ball valves, Reverse Osmosis set-up, and Smart water vending machine. For the control of the plant programmable logic control (PLC) is used, and data acquisition and control are done through supervisory control and data acquisition (SCADA) system.

Only the measurements of the sensors in the considered path are taken for training the neural network. The sensors in consideration are: Pressure Transmitters: 1, 4, 5, 6 and 8, Temperature Transmitters: 1, and 3, Flow Meters: Solar, 5-15, and Vibration sensor.

A detailed information regarding the transducers present in the plant are given in the table below.

TABLE III: TRANSDUCERS AND EQUIPMENTS PRESENT IN THE WATER DISTRIBUTION NETWORK

TDS Indicator	2
Pressure Transmitter	8
Turbidity Analyzer	2
Temperature Transmitter	3
Ultrasonic Smart Flow Meter	14
Ultrasonic Level Transmitter	2
Electromagnetic Flow Meter	1
pH Transmitter with Sensor	2
Vibration Sensor	3
PIV Valve	2
Two-way Solenoid valve	8
Programmable Logic Control	1
SCADA	1
Smart Water Vending Machine	1
Solar Power operated Smart Flow Meter	1
Purified RO Water Storage Tank	1
Pipes and Fitting Materials	
Collector	
Concentrator	
Pumps with safety protection panel	
Electrical cable	
Extension boards	
UPS with Battery Backup	

To collect the data, the plant was run in pump frequency ranging from 30 to 48 Hz, and first, the data for no leak was recorded. The SCADA system automatically records the data taken from the meters in a csv file, with a rate of 6 samples/minute. After the no-leak data record, leaks of different nature and magnitude were created using all the leak points available. The data of the transmitters were again recorded. The final dataset was labelled as leak or no leak so that it can be fed to the supervised learning algorithm. Finally, only the data of relevant transducers were kept and the data was shuffled, to compile the final dataset.

4. NEURAL NETWORK MODEL

After collecting the data, it was pre-processed to keep only the relevant information which contribute to the leak detection classification, as per the studies. Only the measurements of the transducers in the direction of the path of water flow will be considered for the classification model.

The final data set after processing consisted of 866 data samples, with 515 samples of leak and 351 samples of no leak. The input was a 20X1 vector and desired output was a 2-dimensional vector, labelled as leak and no-leak, as one-hot representations.

Resultantly, the input layer of the neural network will be 20-dimensional, and output layer will have 2 neurons. As the 4 primary parameters to classify leakage are pressure, temperature, flow rate and vibration, the penultimate hidden layer should have 4 neurons. To keep the number of parameters to be low, as we have a small dataset, no extra layer was taken. Thus, the network has 1 dense hidden layers 4 neurons respectively and output layer has 2 neurons.

Activation function used in the hidden layer is ReLU, and in the output layer is softmax. Based on the above description, the network has 92 trainable parameters, which results in a good classification model. The network architecture is described as follows:

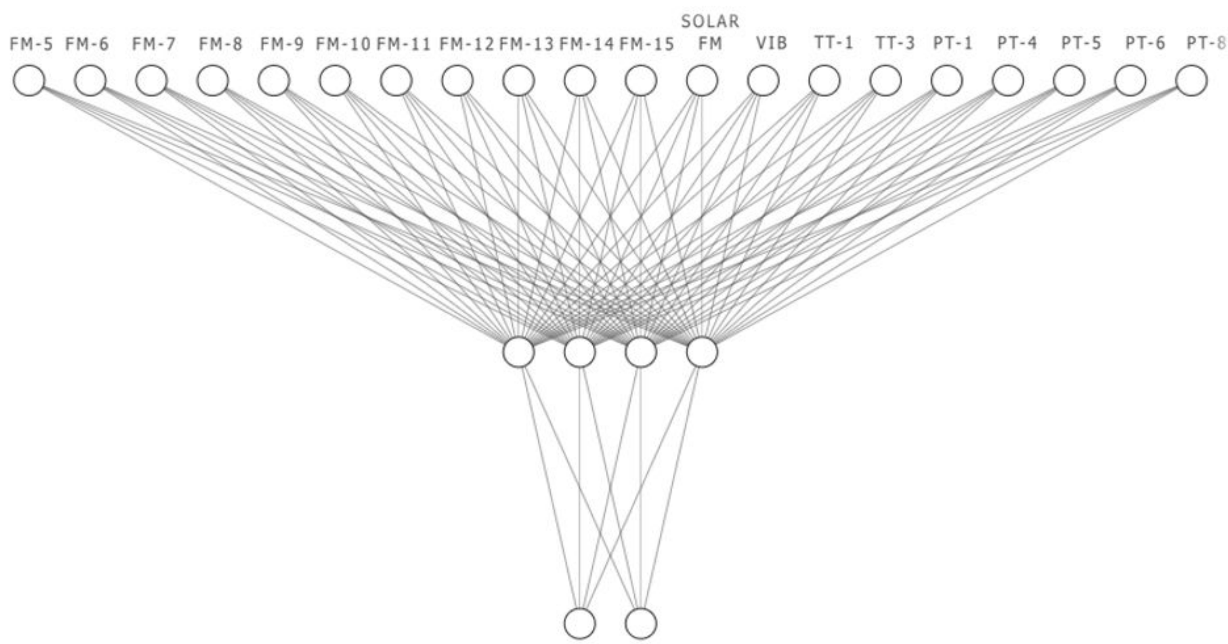

Figure 7: Neural Network model

Before training, the data were shuffled and data was split as 766-100 for training and testing respectively.

The model was trained with this data set for 100 epochs. Loss function was used as 'binary cross-entropy' loss, and 'adagrad' optimizer was used. Adaldelta was preferred because the fluctuations in accuracy were very large with a standard optimizer like stochastic gradient descent, and Adam Optimizer.

Adagrad is adaptive learning optimizer. In the adagrad optimizer, the learning rate is adaptive, in such a way that the learning rate of the parameter gets changed based on how frequently it is modified during backpropagation. This avoids the problem of considering various factors like momentum factor, learning rate, etc. for optimum convergence.

Binary cross-entropy loss is a standard loss function used in a binary classification model, described as:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Binary Cross-Entropy / Log Loss

The activation function ReLU, is as described below:

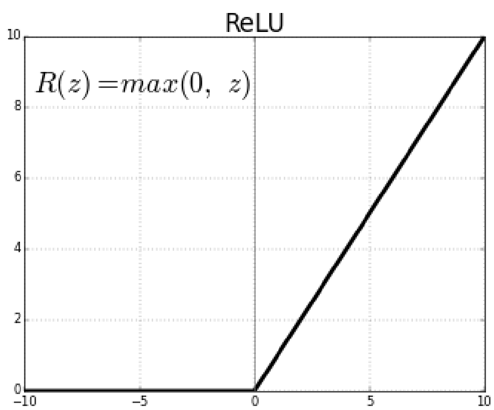

Figure 8: ReLU Activation function

Softmax is a standard activation function used in output layer for a classification problem.

The block diagram of the model, which shows the input and output vector size, is:

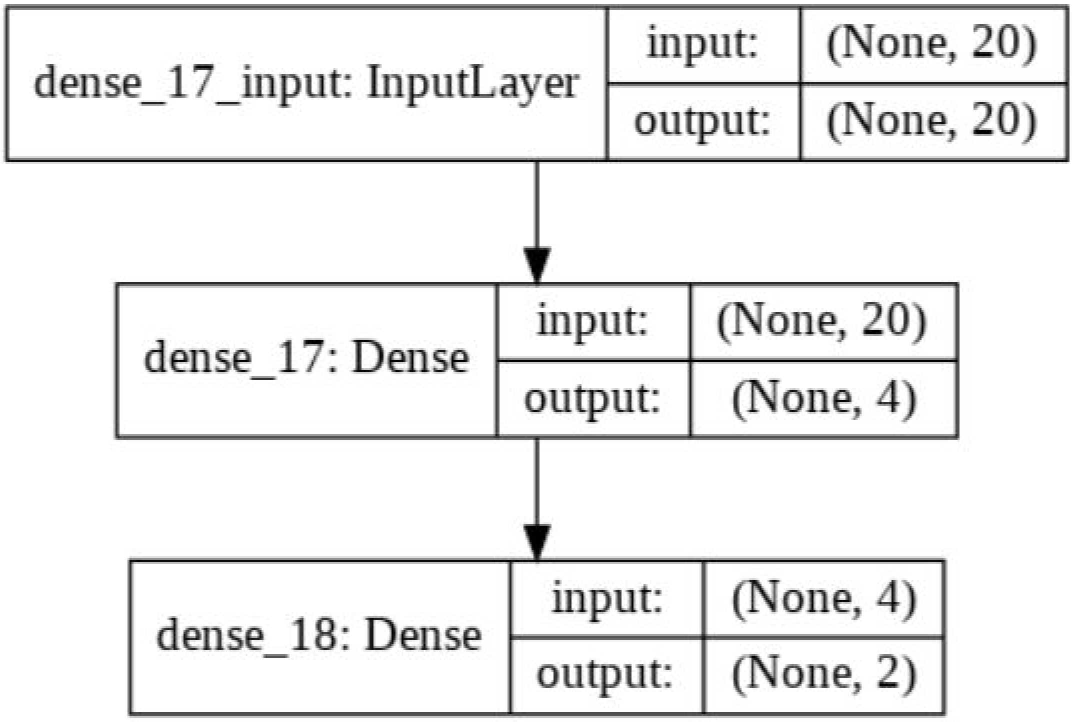

Figure 9: Block diagram of the network

The model was simulated and evaluated in terms of precision, recall and F1 measure.

To evaluate the model quantitatively, the metric precision, recall and F1 measure were used. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations, which shows how good a model is in not generating false alarms. Recall is the ratio of correctly predicted positive observations to all observations in actual class - yes. Hence, recall defines how many, out of all belonging to the class, were correctly predicted to be in it. F1 Score is the harmonic mean of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively, it is not as easy to understand as accuracy, but F1 score is usually more useful than accuracy, especially for uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

5. RESULTS

On simulation after 100 epochs, the graph of loss functions are obtained as follows (on the training data) -

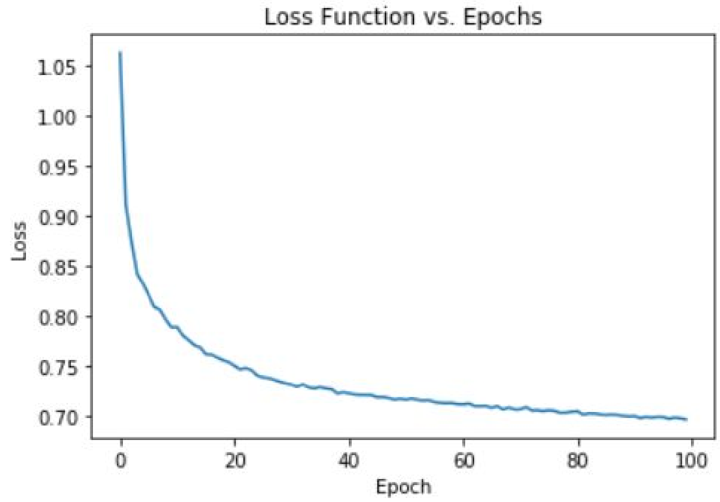

Figure 10: Loss function

This shows that the loss function converges after 100 epochs, and just oscillates around a small value. The loss value after 100 epochs on the training data is obtained to be 0.6. The convergence shows that the model is optimally trained.

With more variant data in hand, this model can be improved. However, simply taking more data won't help the model because the dataset used will only contain redundant data (because it contains data of all possible operating conditions). So, varying more parameters like temperature, TDS, etc. will increase the number of data samples. This modified dataset will result in a better model.

Further, on testing the model on entire data, the confusion matrix is:

On using the model in which the data for testing was separated initially, the confusion matrix, on 100 separated data samples comes out to be as given below:

TABLE IV: CONFUSION MATRIX OF 100 SEPERATE DATA SAMPLES

	Predicted Leak	**Predicted No Leak**
Actual Leak	58	3
Actual No Leak	6	33

This shows that the proposed model works well not only for the originally supplied data, but also for a set of data which has not been encountered before. Thus, it can be practically used for any sort of leak detection, after training done through a large sample of data.

On evaluating this for both the above described, individually, the values that were obtained (for leak detection) proves the above assertion. The values are:

TABLE V: MODEL EVALUATION PARAMETERS

MEASURE	ON COMPLETE DATA
PRECISION	90.6%
RECALL	95.1%
F1 MEASURE	92.8%

6. CONCLUSION

The primary goal of this model is to accurately detect leaks in water distribution system. If this is seen in terms of deep learning, it clearly means that the recall of leak detection should be prioritized as compared to precision, because recall means to successfully classify a leak as one. If the recall is low, the loss will be very high, because the leak will not be identified. Proper remedies would not be taken and as a result, there will be massive loss of resource and money.

Precision, although secondary, is very important, as it determines false alarms. If a no-leak condition is classified as a leak, there will be no loss of resources as such. However, amendment measures will be initiated when it was not required, and it will result in higher operational costs, thus, still leading to a loss.

This model is capable of handling both these losses to a very large extent, because its precision and recall, both are high.

Out of all the time the leak will occur, 95.1% of the time, it will be classified as a leak, and thus, alarm will be raised.

Out of all the times that the alarm will be raised, 90.6% of the time, it will be true. Thus, only 9.4% of alarms will be false.

The overall measure of the model, i.e., its F1 score is 92.8%. This shows that the model balances both the factors to a good extent.

7. SCOPES FOR FUTURE

The project taken is new and has a vast scope in the future. The model developed through this project is based on the output of the sensor which act as features to detect a leak in the pipeline network. This is motivated from the multiple model algorithm. However, transient analysis method can also be extended, and leak can be predicted from the transient change of parameters. This can help in predicting the leak before it occurs, thus minimizing loss to a much larger extent.

Recurrent Neural Network, particularly those made using long short-term memory (LSTM) blocks, can be used to study the time-series data of development of leaks. The output of these RNNs can be further sent to dense layers for classification. The classification is still done by the dense network, but the features are developed using the RNNs, based on sequential data.

However, one constraint is that the data acquisition system must have a high sampling rate to record the transient development rate, before it reaches steady state.

One more scope is to introduce variations in the data, as it was earlier asserted that more number of data would result in better classification accuracy. Since as of now, the only factor that can be varied is the motor speed, that too in the range of 30-50Hz, the variation in data is very limited, particularly when there is no leak.

Many other samples of data can be collected by varying the temperature of water, TDS, etc. However, it requires additional hardware setups to create such conditions. Various types of leak can also be created to incorporate variations.

There can be various other techniques that can be implemented for leak detection. It is evident that combining traditional techniques with artificial intelligence can lead to better models for leak detection and help in the direction of development of societies and smart cities.

REFERENCES

[1] MOHD ISMIFAIZUL MOHD ISMAIL el Al., A Review of Vibration Detection Methods Using Accelerometer Sensors for Water Pipeline Leakage

[2] José-Roberto Bermúdez et Al., Modeling and Simulation of a Hydraulic Network for Leak Diagnosis

[3] Operation Manual, IOT Based Smart Water Distribution System, CSIR-CEERI, Pilani

[4] Panagiotis Tsakalides et Al., Smart Water Grid: A cyber-physical systems approach

[5] Mohammad Tariq Nasir, Measurement error sensitivity analysis for detecting and locating leak in pipeline using ANN and SVM

[6] The Keras documentation - http://keras.io