Digital Control and Management of Water Supply Infrastructure Using Embedded Systems and Machine Learning

Martin C. Peter*
Department of Computer Engineering, Nile University of Nigeria, Abuja Nigeria
E-mail: nwaukom@yahoo.com
*Corresponding Author

Steve Adeshina
Department of Computer Engineering, Nile University of Nigeria, Abuja Nigeria
E-mail: steve.adeshina@nileuniversity.edu.ng

Olabode Idowu-Bismark
Department of Electrical and Information Engineering, Covenant University Ota, Nigeria
E-mail: idowubismarkolabode@gmail.com

Opeyemi Osanaiye
Department of Mechatronics Engineering, Nile University of Nigeria, Abuja Nigeria
E-mail: opeyemi.osanaiye@nileuniversity.edu.ng
ORCID iD: https://orcid.org/0000-0002-0832-307X

Oluseun Oyeleke
Department of Computer Engineering, Nile University of Nigeria, Abuja Nigeria
E-mail: contactseun@gmail.com

Received: 02 May 2023; Revised: 24 July 2023; Accepted: 11 August 2023; Published: 08 October 2023

Abstract: Water supply infrastructure operational efficiency has a direct impact on the quantity of portable water available to end users. It is commonplace to find water supply infrastructure in a declining operational state in rural and some urban centers in developing countries. Maintenance issues result in unabated wastage and shortage of supply to users. This work proposes a cost-effective solution to the problem of water distribution losses using a Microcontroller-based digital control method and Machine Learning (ML) to forecast and manage portable water production and system maintenance. A fundamental concept of hydrostatic pressure equilibrium was used for the detection and control of leakages from pipeline segments. The results obtained from the analysis of collated data show a linear direct relationship between water distribution loss and production quantity; an inverse relationship between Mean Time Between Failure (MTBF) and yearly failure rates, which are the key problem factors affecting water supply efficiency and availability. Results from the prototype system test show water supply efficiency of 99% as distribution loss was reduced to 1% due to Line Control Unit (LCU) installed on the prototype pipeline. Hydrostatic pressure equilibrium being used as the logic criteria for leak detection and control indeed proved potent for significant efficiency improvement in the water supply infrastructure.


1. Introduction

Water is an essential resource and everyone has access to clean and pollution-free water through the use of pipe distribution networks. These pipework’s are buried underground for several years and thus, due to aging, corrosion issues, and third-party damage, cracks occur along the pipes, which causes the loss of significant volumes of water.
Since they are underground and are not readily accessible, monitoring becomes a serious challenge. In a severe situation, the problem becomes more complex when leaks occur significantly at several points along the pipes resulting in supply shortages and contamination as alluded by authors in [1–4]. In view of this, this research targets to provide a solution by focusing on mechanisms to detect and shut fluid leakage when they occur due to breakage of pipeline using hydrostatic pressure and digital control system methods. Existing industrial scale infrastructure for water, oil and gas supply pipeline implementations researched are mostly reliant on destination endpoint supervision systems as means of detecting comprise in the flow of supply received from the production end. This is common with waterboard, oil and gas companies especially in developing countries like Nigeria. The methodology proposed in this research brings an improvement over the legacy method by facilitating real-time and direct detection and shutdown of flow through any physically compromised or broken pipeline which may be situated in remote location far away from user or destination endpoint of the infrastructure and restores flow automatically when fault point is resolved. Fault incidents are recorded in a progressive database to enable pattern modelling using Machine learning for efficient maintenance planning and infrastructure management. While this proposed method proves potent in its functionality, it does not identify the exact point of breakage along the pipeline segments. Fault point identification is an area being considered in future research, as an evolution of the methodology proposed in the current research work.

This paper is structured into five sections beginning with Section 1 covering introduction, Section 2 covers the theoretical background of the research. The method used in actualizing the background principles is covered in Section 3. Experimentation results discussions and conclusion are covered in sections 4 and 5 respectively.

2. Theoretical Background

The theoretical framework of this research is centred on fault detection and control, as the core objective of this work is solving the problem of water supply infrastructure efficiency orchestrated by distribution losses along pipelines. Leveraging Pascal principles and the concepts of hydraulics elaborated by Serrin,[5], a pipeline segment can be analytically represented by indications of various hydrostatic pressure components exerted around the pipe segment by water passing through it as shown in fig.1.

Based on Pascal’s theorem of hydrostatic equilibrium, at a steady state, the sum of hydrostatic pressure components equates to zero as shown in equation (1).

\[ P_{\text{in}} + P_1 + P_2 + P_3 + \ldots + P_n + P_{\text{out}} = 0 \] (1)

Meaning that, \( \sum P_i = 0 \). At a steady state, the liquid water flowing through the pipe exerts undiminished hydrostatic pressure along the interior wall of the pipe while it flows through it. This state is maintained irrespective of the geometric shape taken by the pipe, but this equilibrium state is disrupted once the pipe segment is breached by way of break or rupture. From the prototype schematics shown in fig.3, the pressure components on the pipe section illustrated in fig.4 can be analyzed using the Pascal principle stated in equations (1) to (7). In the analysis below, atmospheric pressure on the water surface in Tanks A and B as shown in fig.2 is represented by \( P_A \) and \( P_B \) in fig.4. This establishes a hydrostatic equilibrium condition equivalent to the pressure equilibrium established by a steady flow of water through an unperturbed pipe. \( P_S \) and \( P_L \) are values of pressure at the sensor and leak points respectively.

\( P_A \) and \( P_B \) represent the sum of atmospheric pressure and pressure due to the column of water in containers A and B respectively. Based on Pascal’s principle, the pressure values at these points of focus are represented in equations (2) and (3)

\[ P_A - P_S - P_L + P_B = 0 \] (2)

\[ P_A + P_B = P_S + P_L \] (3)

When the Leak Valve LV is in a closed position, \( P_L = 0 \) as there is no leak. The sensor pressure value is \( P_S \) at this time, and when in an open position \( P_L \) is not zero, and the sensor pressure is \( P_{S2} \). The value of \( P_{S1} \) is greater than \( P_{S2} \). The differential pressure \( \Delta P \) is the parameter used by a software program in the Embedded systems microcontroller to open or close the control value based on the control algorithm. This is mathematically explained by equations (4) to (7).

\[ P_S = P_A + P_B - P_L \] (4)
As \( P_L \) is Zero due to no leak, \( P_{S1} \) changes to a new value \( P_{S2} \) when is open, \( P_L \) is no longer zero.

\[
P_{S2} = P_A + P_B - P_L
\]

\[
\Delta P_S = P_{S1} - P_{S2} = (P_A + P_B) - (P_A + P_B - P_L)
\]

\[
P_{S1} = P_A + P_B
\]

In Mathematical terms, the value of \( \Delta P_S \) is equal to the hydraulic pressure with which water escapes the pipe wall at the point it was opened via an experimental valve, or a rupture point due to breakage. In terms of control, the sensor threshold is set at a value equal to \( P_{S1} \) so that a drop from that value triggers actuation at the control valve.

Below are the Prototype design conditions considered:

- At steady state, \( P_L = 0 \), Sensor pressure, \( P_{S1} = P_A + P_B \)
- At leak event state, \( P_L \neq 0 \), Sensor pressure, \( P_{S2} = P_A + P_B - P_L \)
- The drop in hydrostatic pressure within the pipe segment is \( \Delta P_S = P_L \)
- The sensor threshold is set at \( P_{S1} \) and drops by \( P_L \) to trigger pipe closure actuation.
- \( P_A \) and \( P_B \) are sums of Atmospheric pressure and pressure of water columns in respective tanks.
  \[ P_A = P_B = P_{atm} + P_{water}. \quad P_{atm} = 101KPa, \quad P_{water} = H \times d \times g = 0.36 \times 1000 \times 10 = 3600Pa. \]

Where \( H \) is the height of water column in the tanks, \( d \) is the density of water and \( g \) is the acceleration due to gravity. Hence, \( P_A = P_B = P_{atm} + P_{water} = 101 + 3.6 = 104.6 \) KPa

\[
P_{S1} = P_A + P_B = 2 \times (104.6) = 209.4 \text{ KPa}
\]

- Sensor maximum threshold value in the control program is set at 200KPa
- The capacity range of pressure sensor used in the prototype is 0 to 5MPa.
Another theoretical aspect reviewed in this research was Infrastructure data which is collected and used for evaluation of the functional efficiency of the infrastructure. Data harvesting from a water supply infrastructure just like other distributed control systems relies on the use of sensors incorporated in data acquisition devices. In a typical pipeline infrastructure using Supervisory Control and Data Acquisition (SCADA) system, Remote Terminal Unit (RTU) can be used for flow data acquisition. Harvested data are then repositioned in a database platform located at the facility’s central office. Such database platforms are computer systems equipped with data analysis software used for data analytics from which insights and knowledge of the infrastructure efficiency as well as its value delivery capacity are determined. Data has become a crucial commodity for decision-making at different levels of infrastructure and enterprise management as alluded by Minh Dang et al.[6] while elaborating the import of data in IoT healthcare application.

Many authors have researched water management problems in the past including the authors of [7] who designed and tested a fuzzy logic-based water pump controller to control the amount of water pumped through a pipe using a flow rate parameter. Patil and Bharkad [8] demonstrated the use of temperature and weight of liquid as parameters for controlling flow rate. Moon and lee [9] proposed the use of a hybrid algorithm of fuzzy and Proportional and Integral (PI) mechanisms to control the temperature of molten glass in the glass manufacturing industry. In the work of Neelshetty et al. [10], the authors researched the use of SCADA and PLC to control recycled water treatment plants. The authors of [11] based their research work on the treatment of unclean water from larger sources such as rivers, lakes, and other water bodies. A proprietary SCADA solution was used here, but this featured additional quality-testing parameters such as turbidity and chlorine content test of the treated water. The work of Shashank et al. [12] provide useful insight into the use of inexpensive hardware such as Microcontroller-based Arduino with open-source software for data acquisition. This serves as a cheaper alternative to a full-scale SCADA deployment. The work of Bakker et al. [13] demonstrated the advantage of model-based control over level-based control, as model-based control provides a more stabilized production flow as compared to level-based control. The work of Rahbalaram et al. [14] on predictive analysis for water management using Machine Learning (ML) and Survival analysis was able to show the efficacy of mining performance data related to system faults using ML algorithms. Ross et al. [15] and van de Ven et al. [16] in their work emphasized the importance of intelligent modeling techniques for determining the quality of water production in a shorter space of time with higher accuracy as compared to the use of traditional laboratory test regimes. Large data samples acquired through plant terminals were fed into an ML-driven software where pre-classified quality grades were used to determine the quality of potable water produced after a pre-determined number of treatment iterations. It provided reliable results in real-time with a high degree of accuracy.

3. Method

3.1. Situation Analysis

The first focus of this work is to have a data-driven analysis of an existing pipeline infrastructure in a chosen location in a bit to study the relationship between production and distribution loss of the water supply infrastructure, and any pattern exhibited by the failure profile. The end goal is to model such patterns to enable event forecasts. Empirical data from utility supply infrastructure can be obtained from a supervisory or monitoring platform. Utility supply companies do have this provision and use different names for such platforms. The platform is used for collating and analyzing alarm data received from field RTUs. This research refers to such platform as a Central Office Monitoring Platform (COMP) and can be a software component in SCADA system which could be hosted on premise or in cloud. It provides the means for managing the field infrastructure based on information derived from harvested data. Five main attributes of data, namely: Volume, Velocity, Veracity, Variety, and Value, in the list posited by Sood et al. [17] are crucial for data reliability and analytics via the COMP. SCADA RTUs connected to a flow meter at each tail-end of the water distribution network collects consumption data which the waterboard company often used to profile and determine consumption rate and indirectly make inferences on infrastructure failure profile from changes in recorded consumption profile. While this indirect approach of fault detection may be serving its purpose, a direct detection mechanism that function within the distribution part of the infrastructure provides a more accurate and reliable data and control method for infrastructure fault management. This is where the use of LCU comes in as proposed by this research. LCU placed at head-end of pipe segments at strategic locations will collect fault data in addition to actuation control action it performs. Field data collected using endpoint RTUs and pipeline LCU provides better information for building a reliable predictive model for future utilization and maintenance forecasts using Machine Learning. In this research, linear regression machine learning algorithm was adopted for descriptive analysis of data, while Random Forest regression was used for predictive analysis. The performance of the prediction model achieved was evaluated using the coefficient of determination or R-square. The field data collected are shown in tables 1 and 2. Fig.7 shows a simplified Wide Area Network (WAN) architecture of the data communication Network over which field data can be backhauled to the COMP. The COMP can be accessed locally via a Local Area Network (LAN) or remotely via a secure Internet connection. For purpose of this research, raw data used for supply and fault situation analysis were obtained in CSV file format from a waterboard facility operator in southern Nigeria. The data was cleaned up and represented for analysis. Result of these analyses are available in section 4.2.
Digital Control and Management of Water Supply Infrastructure using Embedded Systems and Machine Learning

Table 1. Water supply data for a metro area facility

<table>
<thead>
<tr>
<th>Year</th>
<th>PS</th>
<th>TU</th>
<th>Production (Litres)</th>
<th>DL (Litres)</th>
<th>Number of homes</th>
<th>Number of schools</th>
<th>Number of Hospitals</th>
<th>NO</th>
<th>NAPWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>404060</td>
<td>3000231</td>
<td>3749289</td>
<td>749055</td>
<td>80812</td>
<td>143</td>
<td>18</td>
<td>297</td>
<td>81</td>
</tr>
<tr>
<td>2001</td>
<td>412413</td>
<td>3693303</td>
<td>4615626</td>
<td>922323</td>
<td>103103</td>
<td>158</td>
<td>20</td>
<td>315</td>
<td>88</td>
</tr>
<tr>
<td>2002</td>
<td>420767</td>
<td>3741981</td>
<td>4676473</td>
<td>934492</td>
<td>105369</td>
<td>214</td>
<td>21</td>
<td>334</td>
<td>92</td>
</tr>
<tr>
<td>2003</td>
<td>429120</td>
<td>3741981</td>
<td>4676473</td>
<td>934492</td>
<td>105369</td>
<td>214</td>
<td>21</td>
<td>334</td>
<td>92</td>
</tr>
<tr>
<td>2004</td>
<td>437474</td>
<td>3516528</td>
<td>4394658</td>
<td>878129</td>
<td>109369</td>
<td>231</td>
<td>23</td>
<td>354</td>
<td>98</td>
</tr>
<tr>
<td>2005</td>
<td>445827</td>
<td>4272553</td>
<td>5339688</td>
<td>1067135</td>
<td>111457</td>
<td>256</td>
<td>25</td>
<td>376</td>
<td>104</td>
</tr>
<tr>
<td>2006</td>
<td>454181</td>
<td>4161231</td>
<td>5200536</td>
<td>1039305</td>
<td>113545</td>
<td>282</td>
<td>28</td>
<td>422</td>
<td>120</td>
</tr>
<tr>
<td>2007</td>
<td>462534</td>
<td>4328865</td>
<td>5410078</td>
<td>111457</td>
<td>232</td>
<td>33</td>
<td>398</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>470888</td>
<td>5622069</td>
<td>1081213</td>
<td>282</td>
<td>43</td>
<td>447</td>
<td>129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>479241</td>
<td>5155309</td>
<td>4102169</td>
<td>107280</td>
<td>191</td>
<td>27</td>
<td>354</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>487595</td>
<td>5223256</td>
<td>4394658</td>
<td>878129</td>
<td>211</td>
<td>31</td>
<td>376</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>495948</td>
<td>6621455</td>
<td>8271816</td>
<td>111457</td>
<td>232</td>
<td>33</td>
<td>398</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>504302</td>
<td>8151046</td>
<td>10183805</td>
<td>2032759</td>
<td>117722</td>
<td>256</td>
<td>25</td>
<td>422</td>
<td>120</td>
</tr>
<tr>
<td>2013</td>
<td>512655</td>
<td>8456086</td>
<td>12063830</td>
<td>2109019</td>
<td>128164</td>
<td>368</td>
<td>36</td>
<td>474</td>
<td>139</td>
</tr>
<tr>
<td>2014</td>
<td>521009</td>
<td>9655066</td>
<td>12063830</td>
<td>2408764</td>
<td>130252</td>
<td>375</td>
<td>37</td>
<td>503</td>
<td>150</td>
</tr>
<tr>
<td>2015</td>
<td>529362</td>
<td>6887872</td>
<td>1716943</td>
<td>132341</td>
<td>386</td>
<td>75</td>
<td>713</td>
<td>223</td>
<td></td>
</tr>
</tbody>
</table>

PS = Population Size  TU = Total Utilization  NO = Number of organizations
NAPWS = No of Alternative portable water sources  DL = Distribution Loss

Fault data collected across 10 pipe cluster (PCL) areas within the 15 years period is shown in table 2 below:

Table 2. Water supply failure data for a metro area facility

<table>
<thead>
<tr>
<th>Location ID</th>
<th>Total Segment Length (m)</th>
<th>Number of failures per year</th>
<th>Failure rate per 100m per year</th>
<th>Average Age of Pipe Segment (Yr.)</th>
<th>Average Depth from Surface (m)</th>
<th>MTBF (Hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCL021</td>
<td>53777.60</td>
<td>8204</td>
<td>15.25</td>
<td>15</td>
<td>0.72</td>
<td>1.07</td>
</tr>
<tr>
<td>PCL022</td>
<td>53181.92</td>
<td>1620</td>
<td>3.06</td>
<td>12</td>
<td>0.93</td>
<td>5.41</td>
</tr>
<tr>
<td>PCL023</td>
<td>42944.48</td>
<td>689</td>
<td>1.63</td>
<td>13</td>
<td>1.10</td>
<td>12.71</td>
</tr>
<tr>
<td>PCL024</td>
<td>3768.48</td>
<td>128</td>
<td>3.38</td>
<td>9</td>
<td>0.66</td>
<td>68.44</td>
</tr>
<tr>
<td>PCL025</td>
<td>6880.16</td>
<td>581</td>
<td>8.44</td>
<td>7</td>
<td>0.98</td>
<td>15.08</td>
</tr>
<tr>
<td>PCL026</td>
<td>21604.48</td>
<td>954</td>
<td>4.44</td>
<td>13</td>
<td>0.97</td>
<td>9.18</td>
</tr>
<tr>
<td>PCL027</td>
<td>1200.00</td>
<td>157</td>
<td>2.62</td>
<td>11</td>
<td>0.56</td>
<td>55.73</td>
</tr>
<tr>
<td>PCL028</td>
<td>2450.00</td>
<td>321</td>
<td>5.35</td>
<td>11</td>
<td>1.10</td>
<td>27.30</td>
</tr>
<tr>
<td>PCL029</td>
<td>1300.00</td>
<td>170</td>
<td>2.84</td>
<td>12</td>
<td>0.00</td>
<td>51.45</td>
</tr>
<tr>
<td>PCL030</td>
<td>1058.48</td>
<td>139</td>
<td>2.31</td>
<td>10</td>
<td>0.99</td>
<td>63.18</td>
</tr>
</tbody>
</table>

Total  188165.60   12963.00     51.93
3.2. Fault Detection and Resolution Mechanism

This aspect of the research focuses on the method used to realize the LCU functionality, hence an aspect dealing with the design of fault detection and control mechanism. The LCU is achieved using a microcontroller-based embedded system. This research adopts the orifice plate method as the sensing mechanism owing to the nature of the use case in focus and the degree of accuracy required in measuring liquid flow for efficient control leveraging the work of Sood et al [17]. Actuation is based on an electro-mechanical valve sub-system. Each pipeline active junction or node comprises of an Embedded System (ES) unit, the orifice plate meter, and an electro-mechanical valve, all being incorporated together either as a retrofit or manufactured into an integrated unit referred to as the Line Control Unit (LCU) as seen in fig.6. For enterprise-grade deployments, ES units equipped with an enterprise-grade real-time operating system will be a much more suitable solution [18-20]. The Embedded System used in this work is ESP8266 equipped with an Atmel AVR microcontroller, 16 MB flash memory, and 16Mhz processor speed. The ES is then interfaced with a wireless internet modem for network communication with fault management terminals. The typical operation of the LCU stems from the operation of the algorithm programmed into the embedded system. The line control units at the head-ends of the pipeline segments are controlled locally and monitored remotely as they communicate with a back-end central office. The target is for these ES devices to communicate the status of water flow in pipe segments on which they have been mounted. It is envisioned that the entire network footprint for a metropolitan area could be up to a few hundred kilometers, and depending on pipe segments between junctions, several LCUs equaling the number of monitored and controlled segments may be required for full-scale project implementation. A best-fitting and cost-optimized ES unit equipped with the right RTOS and programming is key for efficient operation [21-24]. This is crucial for an LCU functionality.

![Internal schematics of a Line Control Unit (LCU)](image)

The typical operation of an LCU stems from the operation of an algorithm programmed into the ES unit. A flowchart illustrating this program is shown in fig.7. The potency of this control method is evaluated based on empirical data obtained from the experiment conducted with an LCU installed on a pipeline prototype shown in fig.2. The experiment data is presented in table 3.

<table>
<thead>
<tr>
<th>Leak Aperture (%)</th>
<th>Threshold: 200 KPa Response Delay (Min)</th>
<th>Threshold: 150 KPa Response Delay (Min)</th>
<th>Threshold: 100 KPa Response Delay (Min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100000.00</td>
<td>100000.00</td>
<td>100000.00</td>
</tr>
<tr>
<td>25</td>
<td>7.20</td>
<td>7.70</td>
<td>8.10</td>
</tr>
<tr>
<td>50</td>
<td>5.00</td>
<td>6.30</td>
<td>6.60</td>
</tr>
<tr>
<td>75</td>
<td>3.70</td>
<td>5.10</td>
<td>5.90</td>
</tr>
<tr>
<td>100</td>
<td>2.00</td>
<td>3.90</td>
<td>4.10</td>
</tr>
</tbody>
</table>

3.3. Fault Control Experiment and Result

Fault control data of a pipe segment fitted with an LCU was collected in a real-time experiment using the research prototype presented in fig.2. The captured data describes the behavior of a typical pipe segment deployed in an expansive field implementation. Actuation control performance of the LCU is greatly dependent on its pressure sensor sensitivity threshold setting. This by design was set at 200KPa as the ideal threshold value hinging on prototype design conditions stated in section 2 of this work. The threshold was thereafter reduced in steps of 50KPa to observe the behavior of the sensor in terms of response time. It was observed that for different leak aperture sizes of the leak valve...
(LV), different response delays occurred in minutes. These are shown in table 3 below. Analysis of the result obtained is presented in section 4.3.

**Fig. 7. Operation program flow chart of an LCU microcontroller**

### 4. Result Analysis and Discussion

This section presents analysis of the results obtained in this research and offers a comprehensive discussion of the findings made.

#### 4.1. Water Production and Distribution Data Analysis

Fig 8 below shows the result obtained by plotting a linear regression graph for the bi-variate relationship between production and distribution loss based on field data tabulated in table 1. The graph in fig.8 shows a resulting linear relationship between the parameters, meaning that given other independent variables at play, distribution loss will increase in proportion with production quantity. Production quantity is budgeted to match consumption or utilization demand, however, if the causative factors of loss remain in place, distribution losses will vary in direct proportionality with production. This is a descriptive analysis of the production versus distribution loss behavior of the infrastructure under study, to provide insight into magnitude or order of the loss profile. To predict the future utilization profile of the water supply user-end, a Random Forest regressor machine learning algorithm was applied on the dataset presented in table 1. By this, a utilization model was derived to predict future utilization from end-users. The performance of this model was evaluated using the coefficient of determination (or R-square). The R-square value achieved was 0.93 or 93%. Based on the above information, it follows that performance prediction of supply utilization will be 93 per cent accurate with new data. Magnitude of loss profiled in the descriptive analysis in conjunction with utilization prediction capability provide useful insights cost-effective water treatment production and infrastructure maintenance planning.

#### 4.2. Water Supply Infrastructure Fault Data Analysis

The result obtained from the plot of MTBF and yearly failure rate from field data in table 2 is shown in fig.9. This descriptive analysis plot was achieved by use of linear regression machine learning algorithm. An approximate pattern generated by the algorithm is shown in fig.9, taking MTBF as the output or dependent variable and other features of the dataset as input or independent variables. Based on the graph in fig.9, the algorithm estimates by linear regression an inverse relationship between MTBF and failure rate per year. Meaning that going by historical data, as the number of failure rates rises due to incidences along pipelines the Mean Time Between Failures decreases. This bi-variate relationship exists between both parameters given the independent variable features listed on the failure data of table 2.
as the driving factors. To predict the MTBF profile based on these driving factors in the dataset, a Random Forest regressor ML algorithm was used. The prediction performance was evaluated based on coefficient of determination (or R-square) metrics. The value obtained was 0.81 or 81% measure of prediction accuracy, meaning the prediction model will be 81 per cent accurate with new data. This prediction accuracy can be further improved using other ML algorithms in stand-alone mode or as hybrid of ML algorithm ensembled to achieve better prediction performance.

**Fig.8. Distribution loss Vs production quantities fit**

**Fig.9. MTBF vs failure rate data line of fit**

### 4.3 Fault detection and Control Result Analysis

Fig.10 and 11 present an analysis of the data tabulated in table 3. This analysis was done graphically using a python program to obtain the graphs represented below. The graphs obtained from the scenarios of observation show a decaying exponential behavior for the relationship between Leak aperture and Response delay for the given threshold pressure setting. With fine granular values of Leak apertures, the graph would come out as a smooth exponential decay curve, inflexing at a point between 20 and 30 percent of the full leak aperture size. These curves depict that shorter response delays under a normalized value of $0.2 \times 10^{-5}$ minutes occur for Leak aperture sizes beyond 20% of leak calibration used in the design prototype, whereas smaller leak apertures below 20% of leak calibration result in longer delays before the leak is detected by the pressure sensor. Where there is no leak, meaning the leak aperture is 0%, the response delay tends to infinity ideally, meaning that no leak has been detected. This infinite delay condition is arbitrarily depicted by $10^6$ on the data table above for illustrative purposes. Python programming tool used for analysis normalized all values on the Response delay axis to simplify the graph plot.

Interpreting the above result, it is fair to say that response delay of the pressure sensor device equipped on the prototype pipe segment LCU showed that for smaller leak apertures less than 20% of the full leak aperture size, there would be prolonged leak of water before the sensor detects the failure event and trigger the controller to drive valve closure actuation. From table 3, the highest empirical value for observed delay under a leak condition is 8.1 minutes and being that the leak aperture is small under this condition the amount of water loss is minimal prior to fault detection and control. In this case, it was only 17cl of water that was lost out of 15 liters in the supply tank, which is around 1% loss. Despite the empirical limitation of response delays, the system would still perform better than when left to leaks for days, months and perhaps indefinitely due to absence of maintenance intervention and automation of this nature, as is the case with most existing water supply infrastructure in developing cities in Nigeria.
5. Conclusions

This research brings a proposition on the use of cost-effective solution leveraging a mix of embedded systems and Machine learning technologies for water infrastructure fault control, utilization data acquisition, processing, and forecasting for efficient water supply management. These indices are the core objectives and focus of this work. From control system programming, systems design, and dataset modeling using Linear and Random Forest regression ML models, it has been demonstrated that the operations of a water supply infrastructure can be modeled accurately from production to consumption stages, having reduced distribution losses caused by pipeline failures to the barest minimum using digital control schemes implementation. This research work brings a valuable contribution towards solving a key challenge faced in the supply of portable water through pipelines in most developing cities where government budget for infrastructure implementation and maintenance is minimal and investments are particularly low due to low returns and huge wastes of treated water in course of distribution via faulty pipelines. Deployment of such cost-effective solution along pipelines not only curb waste but also enable waterboard companies forecast accurately consumption profile of the user-end, hence determine appropriate production capacity per time. These together will boost availability of supply to end-users, revenue assurance to the production companies and thus encourage more investment into such infrastructure and business sector as consumption grows over time.

Acknowledgment

We acknowledge all contributing authors.

References

Chukwuedo Martin Peter is the Director of Engineering Phase 3 Telecom Nigeria. He holds a bachelor’s degree in Electrical/Electronic Engineering from Abubakar Tafawa Balewa University Nigeria, a master’s degree in Computer Engineering from Nile University Nigeria, a master’s in Business Administration from the University of Roehampton United Kingdom, and has over 15 years of cognate experience in Telecommunication engineering. He is a senior member of the IEEE (SMIEEE) and has worked in different committees in a University of Roehampton United Kingdom, and has over 15 years of cognate experience in Telecommunication engineering. He is a senior member of the IEEE (SMIEEE) and has worked in different committees in a

Olabode Idowu-Bismark is a Senior Lecturer at Covenant University, Ota, Nigeria. He obtained his Ph.D. in Information and Communication Engineering from Covenant University Ota, Nigeria. Olabode has over 30 years of cognate experience in Extra Low Voltage (ELV) and Building services Engineering before joining the academics. He is a member of the Nigerian Society of Engineers (MNSE), a member of IEEE (MIEEE), a member of IET (MIET), and a COREN Registered Engineer. He has published many scientific papers in international peer-reviewed journals and conferences. His research interest is in the area of mobile communication, mmWave, and MIMO communication. He can be contacted at email: idowubismarkolabode@gmail.com
**Opeyemi Osanaiye** is received a bachelor’s degree in electrical engineering from the University of Ilorin, Nigeria, in 2007, a master’s degree in telecommunications engineering from the University of Sunderland, U.K., in 2011, and a Ph.D. degree in electrical engineering from the University of Cape Town, South Africa, in 2016. He is a Senior Lecturer and Head of Department at the Mechatronics Engineering Department, Nile University of Nigeria. His research interests include computer networks, cloud computing, wireless sensor network, fog computing, network security, voice-over-internet protocol technology, and cloud computing security. He is a registered COREN member and a member of IEEE.

**Oluseun. D. Oyeleke** is a lecturer 1 at the Nile University of Nigeria. He is a graduate of Electrical engineering from Bayero University Kano and also has his Meng from the Nigeria Defence Academy in Electronics and communications. His research interest is in the areas of microwave technology, solid-state electronics, and machine learning for telecommunication and health applications. He can be contacted by email at oluseun.oyeleke@nileuniversity.edu.ng, contactseun@gmail.com

**How to cite this paper:** Martin C. Peter, Steve Adeshina, Olabode Idowa-Bismark, Opeyemi Osanaiye, Oluseun Oyeleke. "Digital Control and Management of Water Supply Infrastructure Using Embedded Systems and Machine Learning", International Journal of Intelligent Systems and Applications(IJISA), Vol.15, No.5, pp.1-11, 2023. DOI:10.5815/ijisa.2023.05.01